Technical Report

Department of Computer Science & Engineering
University of Minnesota
4-192 Keller Hall
200 Union Street SE
Minneapolis, MN 55455-0159 USA

TR 21-001

Source Aware Modulation for leveraging limited data from heterogeneous sources

Xiang Li, Ankush Khandelwal, Rahul Ghosh, Arvind Renganathan, Jared Willard, Shaoming Xu, Xiaowei Jia, Lele Shu, Victor Teng, Michael Steinbach, John Nieber, Christopher Duffy, Vipin Kumar

August 10, 2021

Source Aware Modulation for leveraging limited data from heterogeneous sources

Xiang Li[2], Ankush Khandelwal[1], Rahul Ghosh[1], Arvind Renganathan[1], Jared Willard[1], Shaoming Xu[1], Xiaowei Jia[1], Lele Shu[4], Victor Teng[2]
Michael Steinbach[1], John Nieber[2], Christopher Duffy[3], Vipin Kumar[1]

[1]University of Minnesota, Department of Computer Science and Engineering

[2]University of Minnesota, Department of Bioproducts and Biosystems Engineering

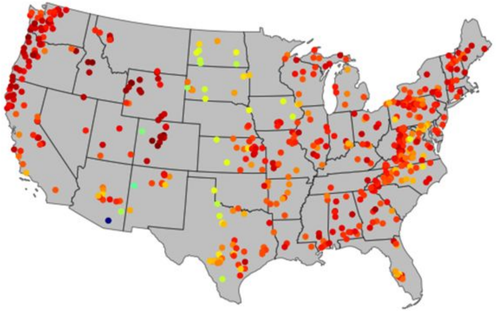
[3]Pennsylvania State University, Department of Civil and Environmental Engineering

[4]University of California-Davis, Department of Land, Air and Water Resources

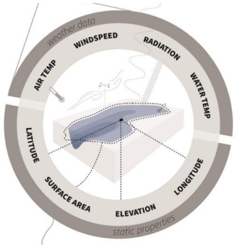
2nd Workshop on Knowledge Guided Machine Learning (KGML2021)

Introduction

The need to combine limited data from multiple **sources** arises in many applications



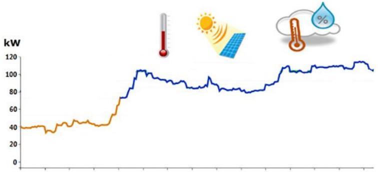
Streamflow prediction using
data from multiple **basins**



Predict lake surface temperature
using data from multiple **lakes**

App usage prediction
using data from multiple **users**

Mood and wellbeing prediction
using data from multiple **patients**



Electrical load forecasting
using data from multiple **households**

Optimal store placement
using data from multiple **cities**

Challenge: Source Heterogeneity

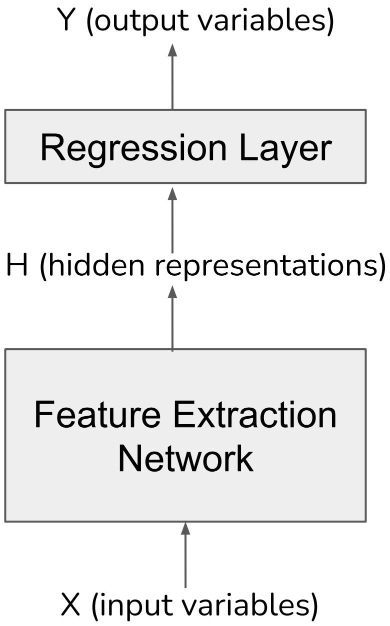
Source characteristics introduce heterogeneity in the relationship between input and output variables

- Impact of weather on mood depends on personality of the **person**[1]
- Impact of weather drivers on streamflow depends on **basin** characteristics[2]
- Impact of demographic features on store placement depends on consumption habits of a **city**[3]

1. Taylor, Sara, Natasha Jaques, Ehimwenma Nosakhare, Akane Sano, and Rosalind Picard. "Personalized multitask learning for predicting tomorrow's mood, stress, and health." *IEEE Transactions on Affective Computing* 11, no. 2 (2017): 200-213.
2. Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, Hydrol. Earth Syst. Sci., 23, 5089–5110, https://doi.org/10.5194/hess-23-5089-2019, 2019.
3. Liu, Yan, Bin Guo, Daqing Zhang, Djamal Zeghlache, Jingmin Chen, Sizhe Zhang, Dan Zhou, Xinlei Shi, and Zhiwen Yu. "MetaStore: A Task-adaptative Meta-learning Model for Optimal Store Placement with Multi-city Knowledge Transfer." *ACM Transactions on Intelligent Systems and Technology (TIST)* 12, no. 3 (2021): 1-23.

Approach: Source Aware Modulation

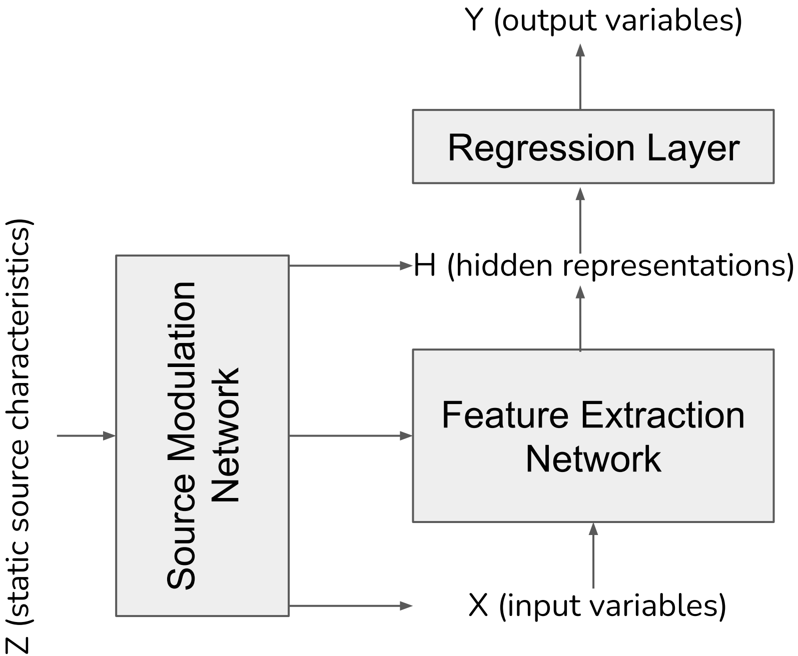
- Learns a single global model but uses source characteristics to modulate the network



Approach: Source Aware Modulation

- Learns a single global model but uses source characteristics to modulate the network

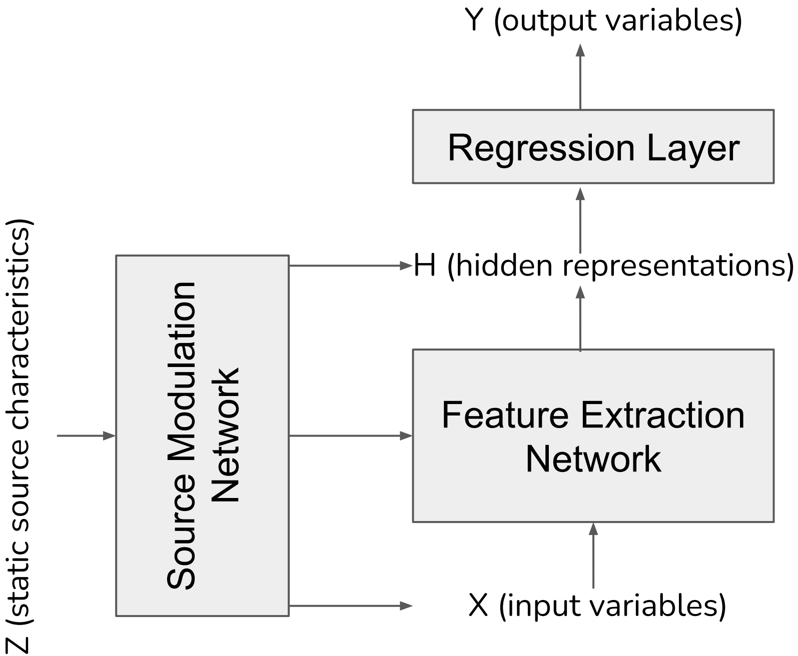
- Modulation can be applied to different parts of the network



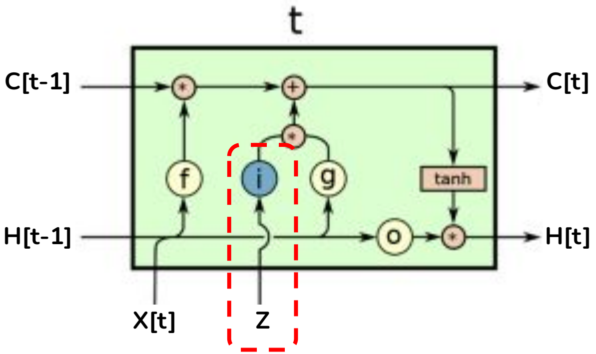
Approach: Source Aware Modulation

- Learns a single global model but uses source characteristics to modulate the network

- Modulation can be applied to different parts of the network



X[t] - dynamic weather variables at time t
H[t] - dynamic hidden representations at time t
Z- static source characteristics
C[t] - dynamic cell state at time t

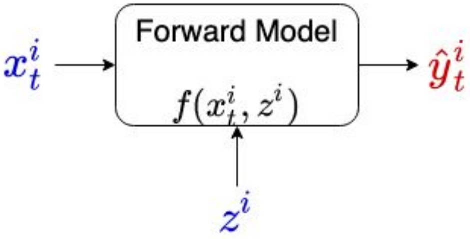


EA-LSTM[1] uses input gate as a modulator driven by static source characteristics to modulate the dynamic cell state

1. Kratzert, Frederik, Daniel Klotz, Guy Shalev, Günter Klambauer, Sepp Hochreiter, and Grey Nearing. "Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets." *Hydrology and Earth System Sciences* 23, no. 12 (2019): 5089-5110.

Outline

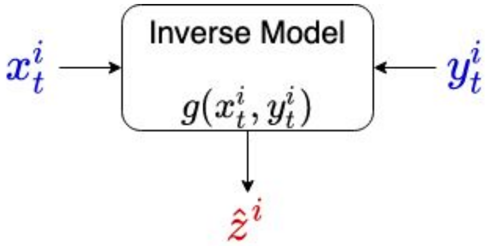
- Randomly assigned characteristics instead of known physical characteristics for source modulation



where i = 1 to N (number of sources)

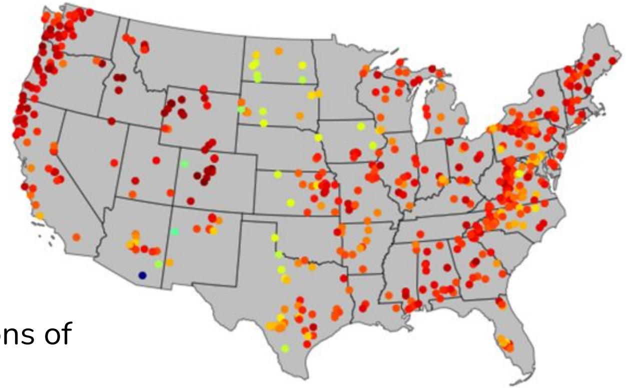
- Reduce uncertainty in static characteristics

- Impute static characteristics



- Identify unknown static characteristics

Experimental Setup

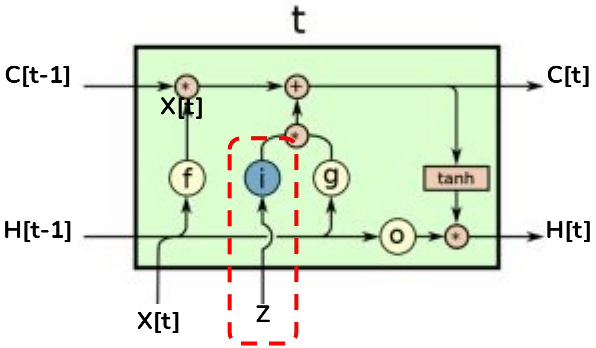


- *Sources:* 531 basins cross the continental US from **CAMELS** dataset
- *Dynamic Inputs: 5 weather variables*
- Dynamic Outputs: streamflow
- *Modulation Strategy:* EA-LSTM
- *Training:* 10 years of data (1999 to 2008) for each basin
- *Testing:* 10 year of data (1989 to 1999) for each basin
- *Ensemble Strategy:* mean prediction from 5 models with different initializations of deep learning model weights
- *Evaluation Metric:* NSE
  - higher values represent better performance

X[t] - dynamic weather variables at time t
H[t] - hidden representations at time t
Z- static source characteristics
C[t] - cell state at time t

$$NSE = 1 - \frac{\sum_{t=1}^{T}(Q_m^t - Q_o^t)^2}{\sum_{t=1}^{T}(Q_o^t - \overline{Q_o})^2}$$

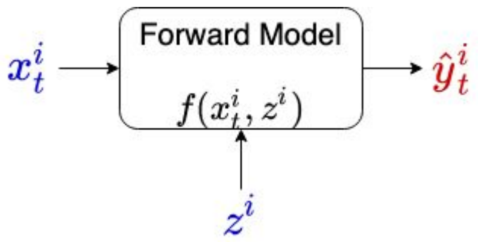
Q_m^t	Predicted discharge
Q_o^t	Observed discharge
$\overline{Q_o}$	Mean of observed discharge



Addor, N., Newman, A.J., Mizukami, N., Clark, M.P., 2017. The CAMELS data set: Catchment attributes and meteorology for large-sample studies. Hydrol. Earth Syst. Sci. 21, 5293–5313. https://doi.org/10.5194/hess-21-5293-2017

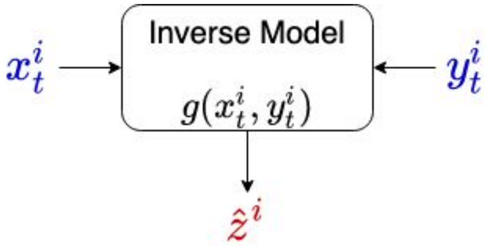
Outline

- **Randomly assigned characteristics instead of known physical characteristics for source modulation**

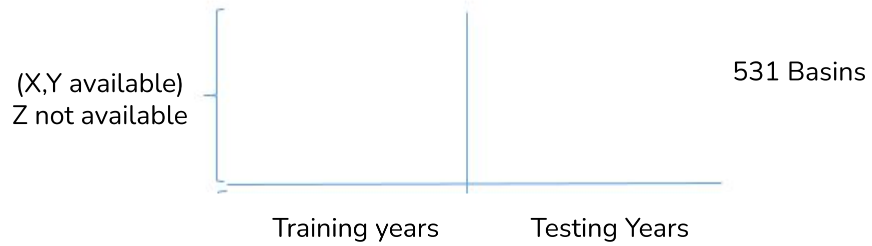


where i = 1 to N (number of sources)

- Reduce uncertainty in static characteristics
- Impute static characteristics
- Identify unknown static characteristics

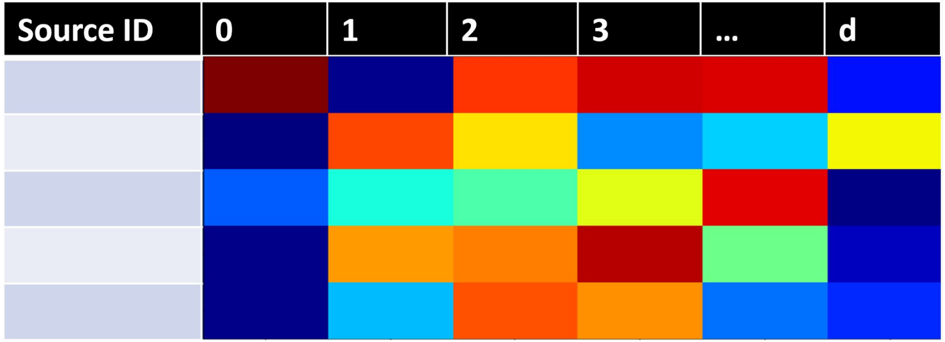


Can we perform source modulation when characteristics are not explicitly known?

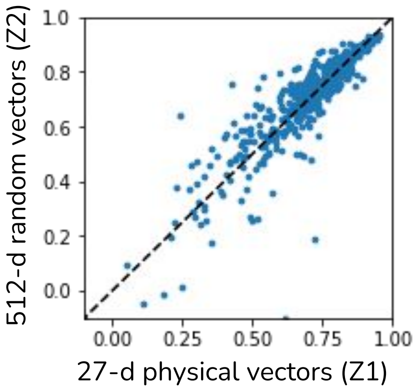


Can we perform source modulation when characteristics are not known?

Random d-dimensional uncorrelated gaussian vectors



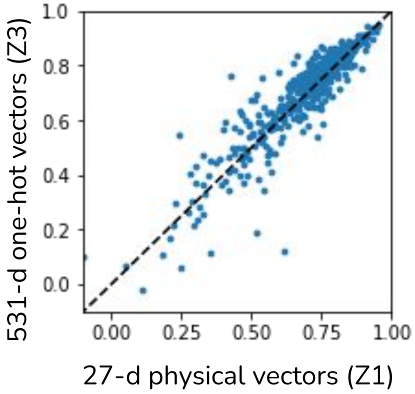
- Maps each source randomly to a point in a *d*-dimensional space
- The choice of *d* will depend on the number and type of sources



One-hot vectors

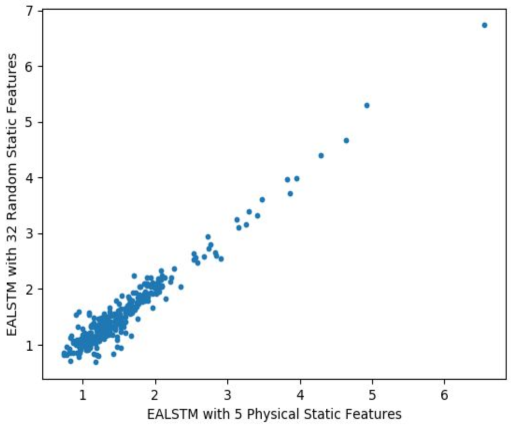
Source ID					...	
	1	0	0	0	...	0
	0	1	0	0	...	0
	0	0	1	0	...	0
	0	0	0	1	...	0
...	...	...	...	...	...	...
	0	0	0	0	...	1

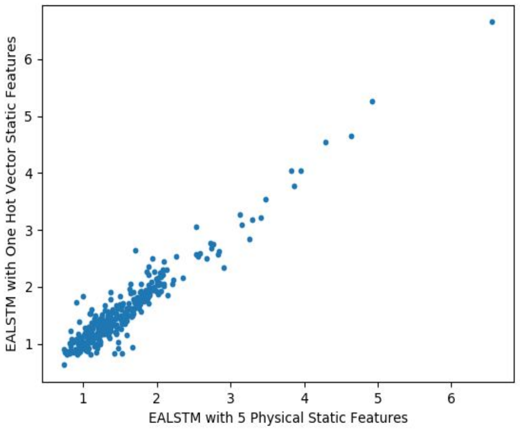
- Gives an index/id to each basin
- The dimension of vectors is equal to the number of sources.
- Assumes every basin equally similar/dissimilar to each other.



Performance of random vectors on LAKE dataset

- Experimental setup
  - *Modulation Strategy:* EA-LSTM
    - 350 days input sequence, 256 hidden units.
  - *Sources:* 345 lakes across Midwestern United States
  - *Dynamic Input and Output:*
  - *Training:* 70 most recent temperature measurements between 1980-2020 for each basin
  - *Testing:* 30 least recent temperature measurements between 1980-2020 for each basin
  - *Ensemble Strategy:* mean prediction from 5 models with different initializations
  - *Static lake Characteristics*
    - 5-d physical vectors - depth, area, latitude, longitude, elevation
    - 512-d random vectors
    - 32-d one-hot vectors
  - *Evaluation Metric:* Root Mean Squared Error
    - lower values represent better performance





Are random vectors robust under different settings ?

- Data sparsity
    - Number of sources
    - Number of observations available for training

- Modulation Strategy
    - EA-LSTM: uses input gate driven by basin characteristics
    - CT-LSTM: concatenates basin characteristics to dynamic inputs
    - FM-LSTM: modulates hidden features from a vanilla LSTM

Are random vectors robust under different settings ?

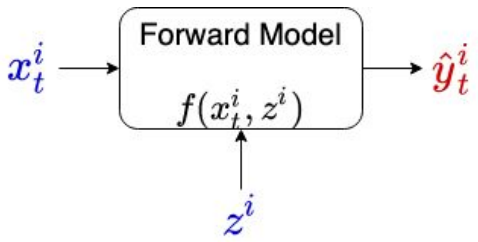
- Data sparsity
  - Number of sources
  - Number of observations available for training

- Modulation Strategy
  - EA-LSTM: uses input gate driven by basin characteristics
  - CT-LSTM: concatenates basin characteristics to dynamic inputs
  - FM-LSTM: modulates hidden features from a vanilla LSTM

- Random vectors show comparable performance to physical characteristics
  - even when only 10 basins are used
  - even when only 10% of training years are used
  - irrespective of modulation strategy

Outline

- Randomly assigned characteristics instead of known physical characteristics for source modulation

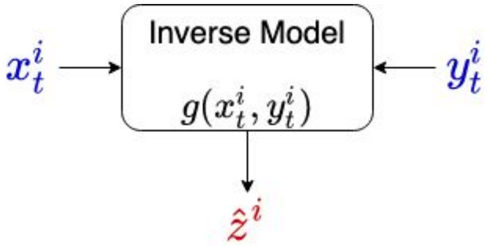


where i = 1 to N (number of sources)

- Reduce uncertainty in static characteristics

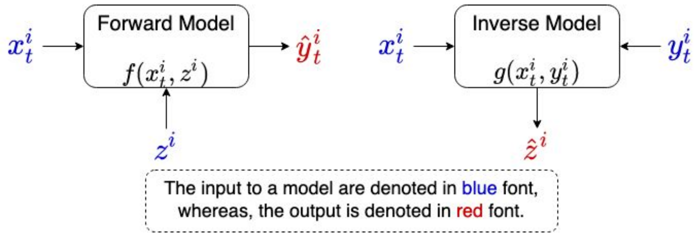
- Impute static characteristics

- Identify unknown static characteristics



Inverse Modeling

- General form of source aware model: $y^i_t = f(x^i_t, z^i)$
  - For streamflow monitoring, y^i_t is the streamflow, x^i_t are the weather drivers and z^i are the time-invariant basin characteristics of basin i



- General form of inverse model: $z^i = g(x^i_t, y^i_t)$
  - captures the time-invariant basin characteristics using time-varying streamflow and weather drivers

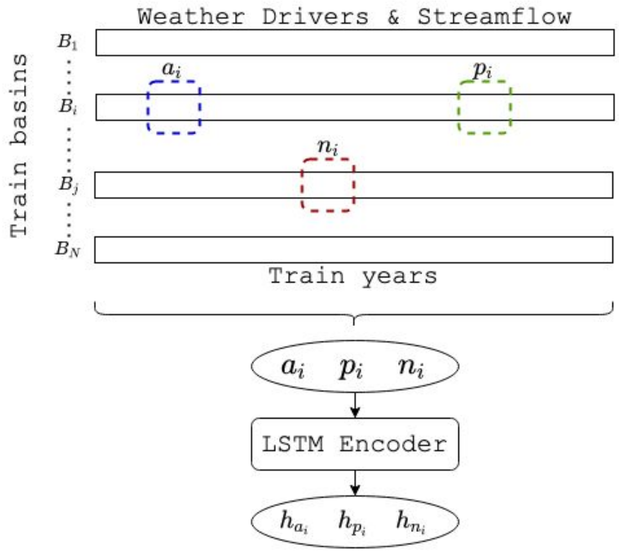
- The inverse model has three key implications:
  - Reduce uncertainty in static characteristics
    - Assumption: characteristics are available but uncertain
  - Impute static characteristics
    - Assumption: characteristics for some basins are missing
  - Identify unknown static characteristics
    - Assumption: no characteristics available for any basin

1. Beven, Keith. "Deep learning, hydrological processes and the uniqueness of place." *Hydrological Processes* 34, no. 16 (2020): 3608-3613.

Approach: Representation Learning

- Learns a deep neural network that generates time invariant and source specific **embeddings** from time varying weather and streamflow data

- Key Idea: The distance between embeddings of any two timeseries samples from the same basin should be smaller compared to timeseries samples from two different basins



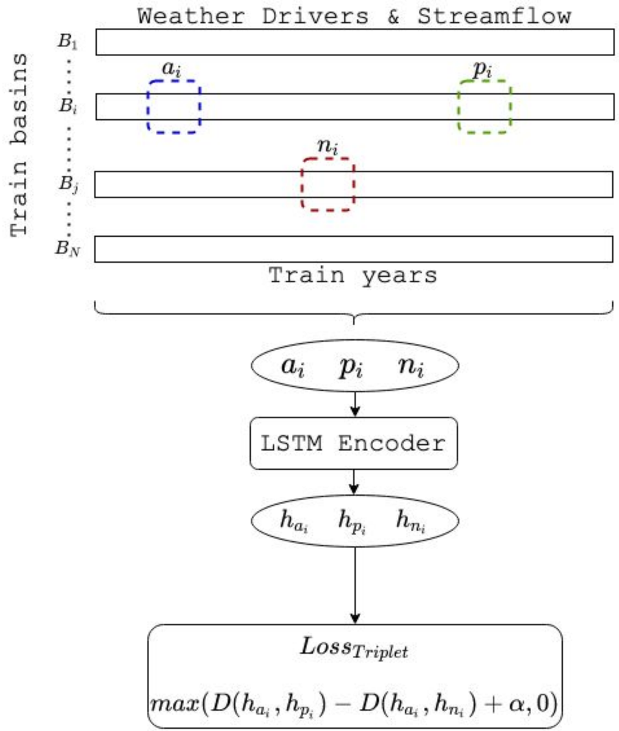
$$\mathcal{D}(h_{a_i}, h_{p_i}) \downarrow \quad \mathcal{D}(h_{a_i}, h_{n_i}) \uparrow$$

Approach: Representation Learning

- Learns a deep neural network that generates time invariant and source specific **embeddings** from time varying weather and streamflow data

- Key Idea: The distance between embeddings of any two timeseries samples from the same basin should be smaller compared to timeseries samples from two different basins

- Triplet loss/Contrastive loss are a widely used loss functions to capture this idea



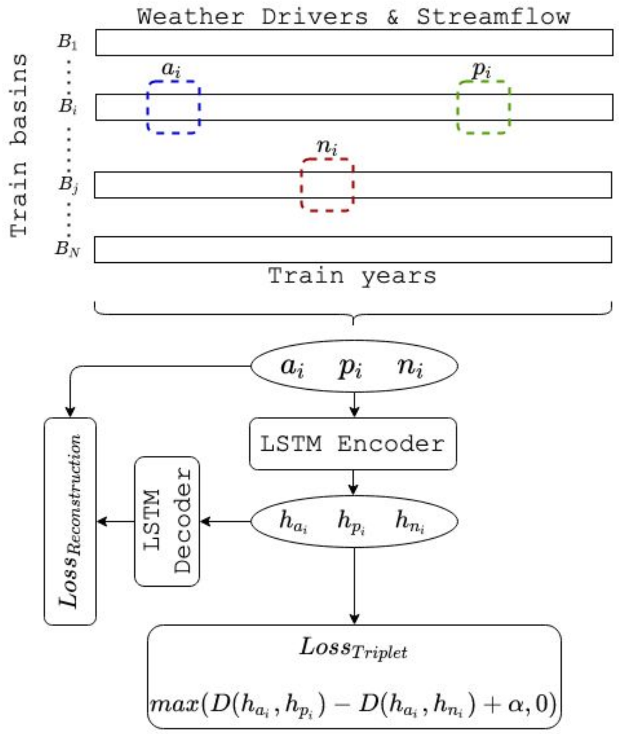
Approach: Representation Learning

- Learns a deep neural network that generates time invariant and source specific **embeddings** from time varying weather and streamflow data

- Key Idea: The distance between embeddings of any two timeseries samples from the same basin should be smaller compared to timeseries samples from two different basins

- Triplet loss/Contrastive loss are a widely used loss functions to capture this idea

- These embeddings can also be regularized by adding reconstruction loss



Approach: Representation Learning

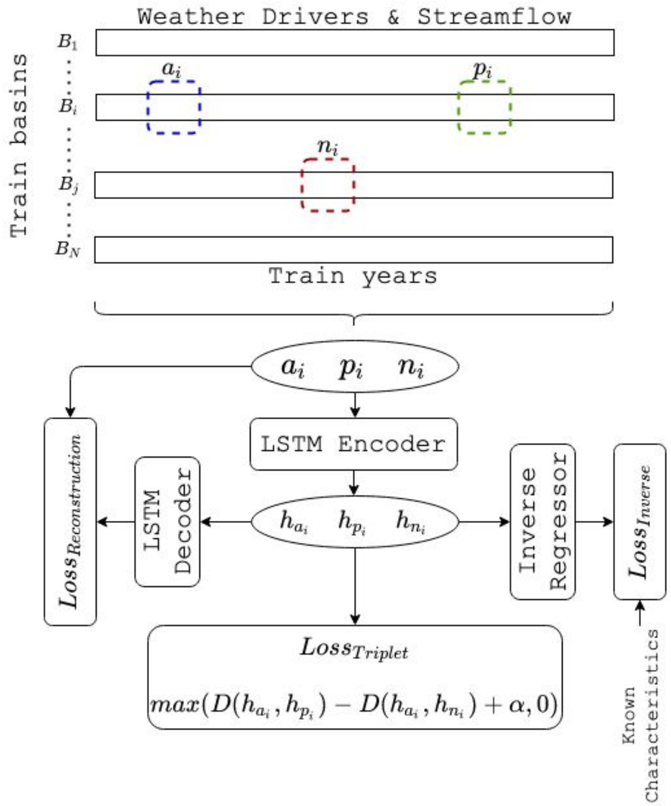
- Learns a deep neural network that generates time invariant and source specific **embeddings** from time varying weather and streamflow data

- Key Idea: The distance between embeddings of any two timeseries samples from the same basin should be smaller compared to timeseries samples from two different basins

- Triplet loss/Contrastive loss are a widely used loss functions to capture this idea

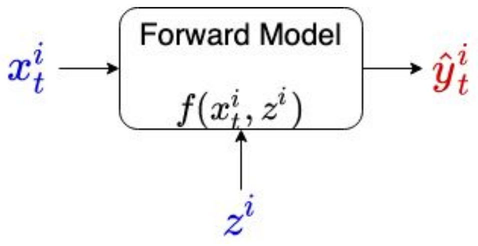
- These embeddings can also be regularized by adding reconstruction loss

- In scenarios where physical characteristics are known (could be uncertain and incomplete)[1], a downstream task to generate known characteristics from learned embeddings can be added as another constraint



Outline

- Randomly assigned characteristics instead of known physical characteristics for source modulation

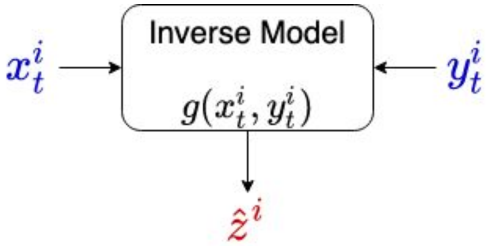


where i = 1 to N (number of sources)

- **Reduce uncertainty in static characteristics**

- Impute static characteristics

- Identify unknown static characteristics

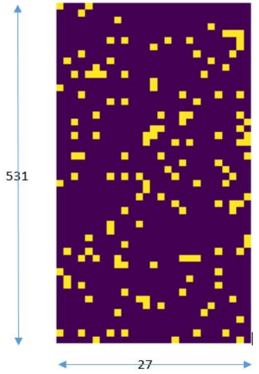


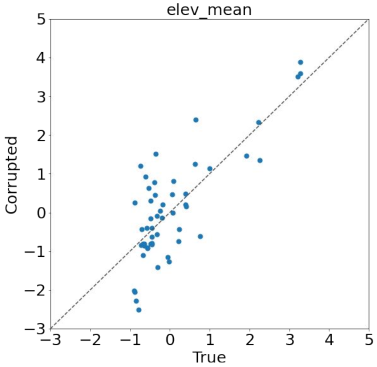
Reduce uncertainty in basin characteristics using **representation learning**

- Measurement uncertainty is very common in hydrological applications
- The inverse model learns generalizable patterns and hence can potentially denoise the corrupted characteristics

Experimental Setup

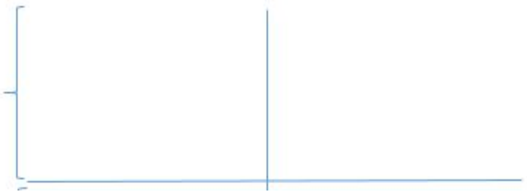
- Add random noise to available characteristics from CAMELS dataset
  - 1 std. deviation noise in 10 % values
- Basin characteristics
  - Z: true characteristics - reference
  - Z̃: corrupted characteristics
  - Z̃rec: reconstructed characteristics using Z̃ during training
  - Zrec: reconstructed characteristics using Z during training
- The model uses unseen test years to reconstruct characteristics to ensure generalizability



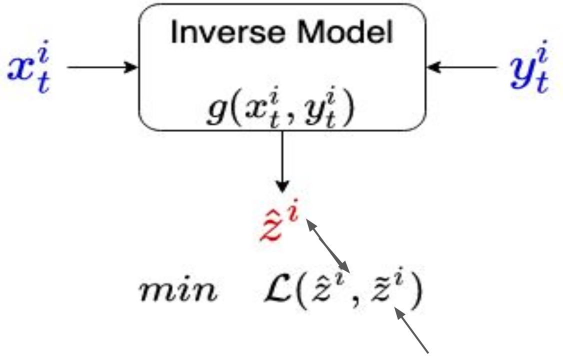


RMSE of corrupted values (Z̃) with respect to true values (Z): 0.83

(X,Y,Z available)
(but Z is corrupted during training)



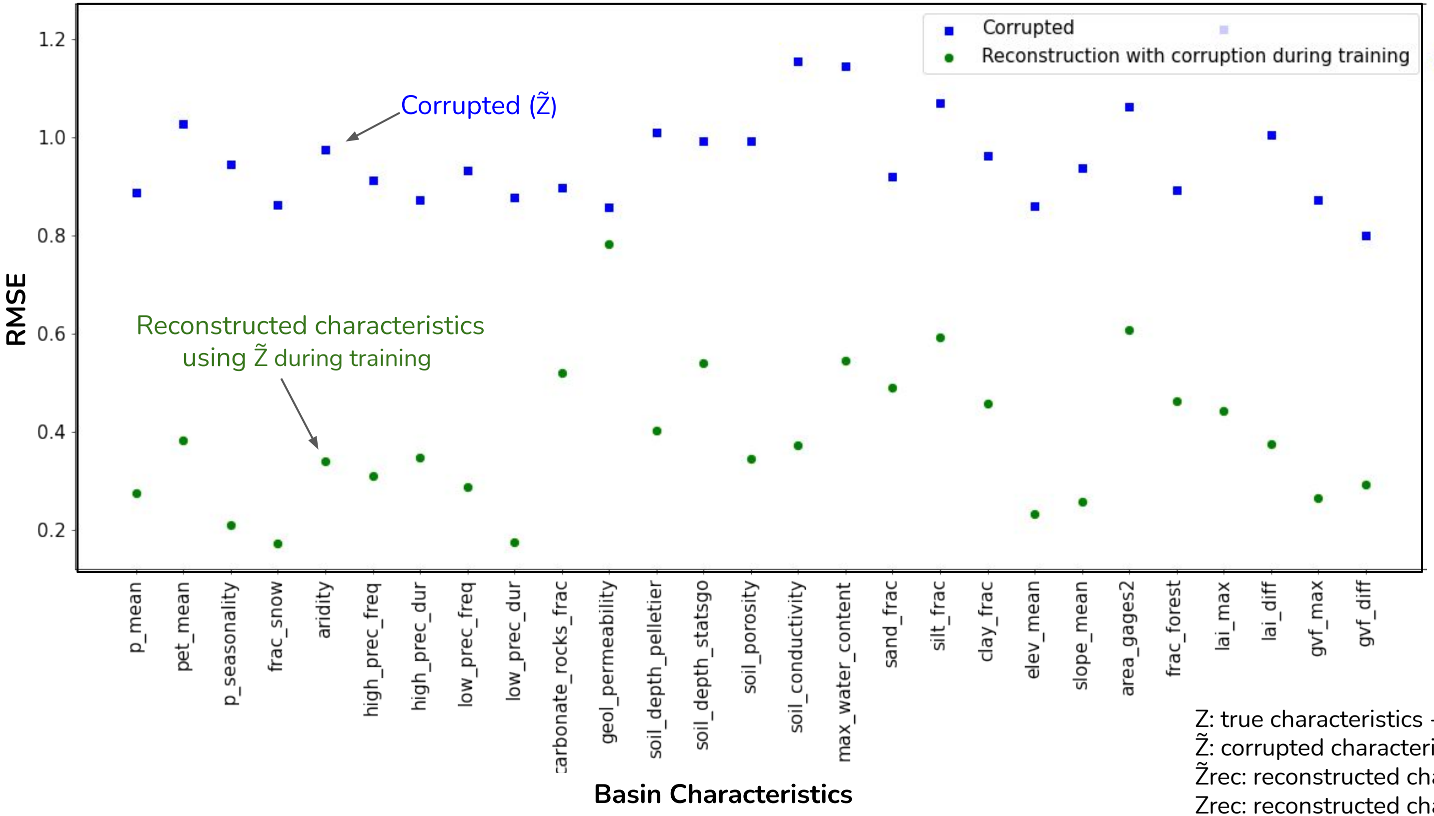
Training years Testing Years

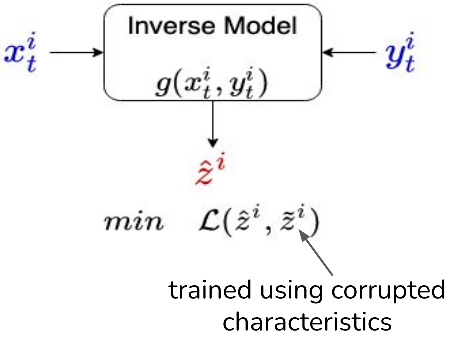


Representation Learning model trained using corrupted characteristics

Impact of measurement uncertainty in basin characteristics

- Representation Learning model significantly reduces measurement error in corrupted characteristics.

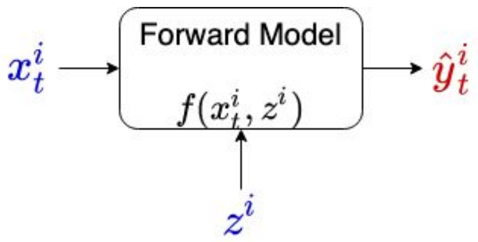




Z: true characteristics - reference
Z̃: corrupted characteristics
Z̃rec: reconstructed characteristics using Z̃ during training
Zrec: reconstructed characteristics using Z during training

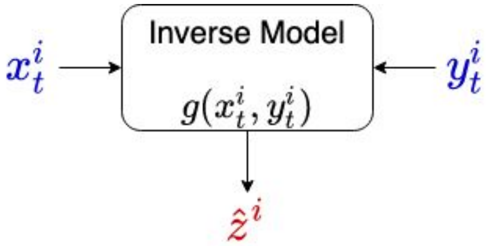
Outline

- Randomly assigned characteristics instead of known physical characteristics for source modulation



where i = 1 to N (number of sources)

- Reduce uncertainty in static characteristics
- **Impute static characteristics**
- Identify unknown static characteristics



Impute missing characteristics

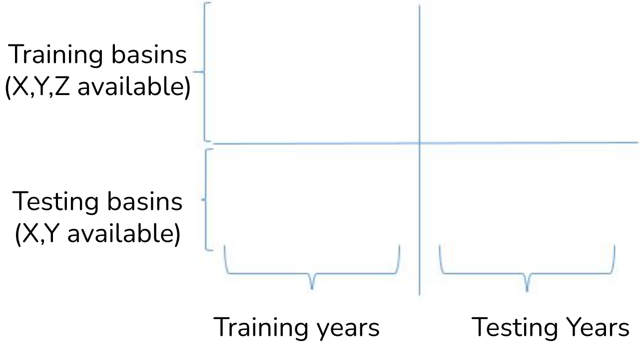
- The inverse model trained on multiple basins can potentially estimate characteristics when they are missing for some basins.

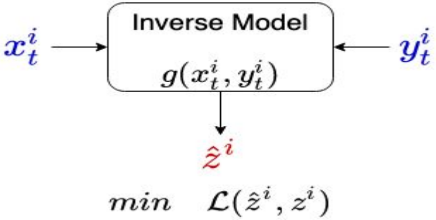
Experimental Setup

- Split the 531 available basins in the CAMELS dataset into two groups, train basins, and test basins.

- We train our model on training basins during train years and predict on test basins during test years.

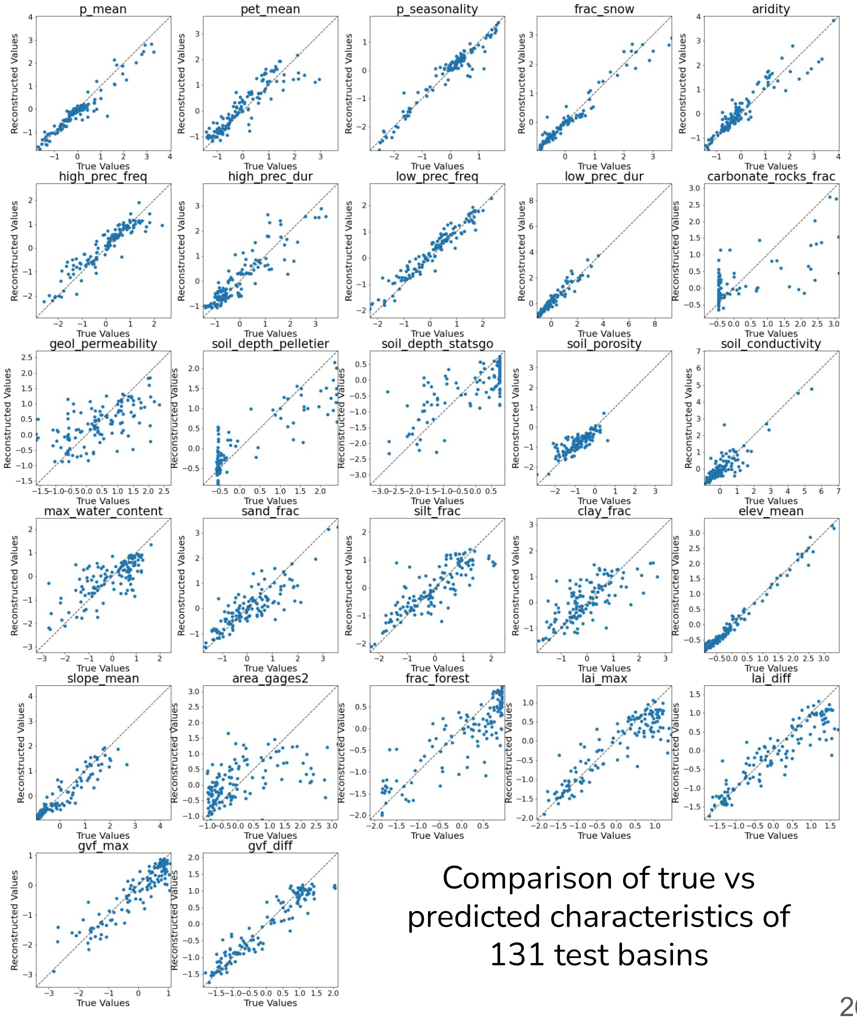
- To evaluate our predictions, we compare the predicteds value with the true values for all the test basins.





Impute missing characteristics

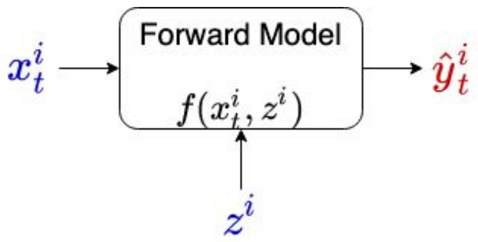
- For many characteristics, the model shows good performance
- Possible reasons for poor predictions for some characteristics
  - Cannot be uniquely identified by weather-streamflow relationship (equifinality)
  - ML architecture is limited
  - High measurement uncertainty



Comparison of true vs predicted characteristics of 131 test basins

Outline

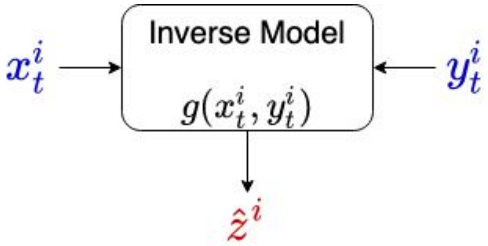
- Randomly assigned characteristics instead of known physical characteristics for source modulation



where i = 1 to N (number of sources)

- Reduce uncertainty in static characteristics

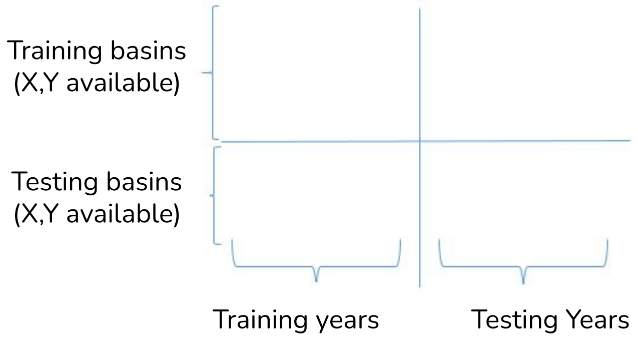
- Impute static characteristics

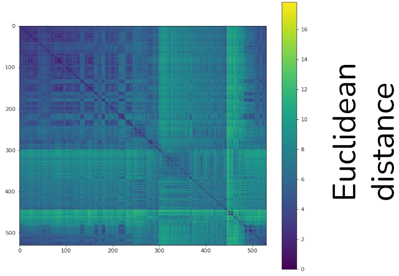


- **Identify unknown static characteristics**

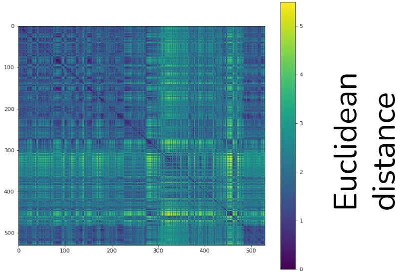
Identify unknown characteristics using self supervised representation learning

- Representation Learning has the potential to identity some time invariant characteristics that may be missing from available characteristics
- We train the inverse model without using any knowledge of available characteristics as a constraint
- Triplet and reconstruction loss are used for model training

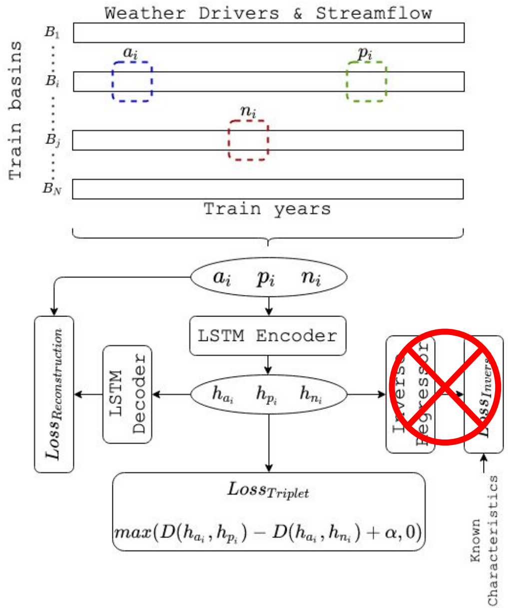




27-d physical vectors



32-d learned embeddings



- Representation learning model generates embeddings that contains meaningful similarity structure between basins

Comparison of learned embeddings with know characteristics

- High correlation between an embedding attribute and known characteristic show ability of the model to learn meaningful representations.

Comparison of learned embeddings with known characteristics

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
p_mean	0.09	-0.26	0.05	-0.40	-0.19	0.06	0.33	-0.39	0.14	-0.14	0.11	-0.05	0.33	-0.01	-0.01	-0.09	0.04	-0.21	0.20	-0.56	-0.10	-0.07	0.46	-0.15	0.08	-0.37	-0.23	-0.24	-0.22	-0.36	0.01	-0.27
pet_mean	-0.40	-0.40	0.42	0.54	0.31	-0.21	-0.68	0.62	0.50	0.26	0.43	-0.23	-0.59	-0.50	0.46	-0.54	0.65	0.28	-0.36	0.24	-0.32	0.70	-0.73	-0.58	0.65	0.08	-0.58	0.29	0.14	0.39	0.46	0.02
p_seasonality	-0.20	0.14	-0.44	0.44	0.57	-0.52	0.11	-0.17	0.12	-0.53	0.17	-0.48	-0.39	-0.31	0.19	-0.15	0.04	0.56	0.19	0.47	0.66	0.12	0.12	0.36	0.07	-0.03	0.26	0.60	0.19	0.54	0.15	-0.32
frac_snow	0.75	0.77	-0.23	-0.62	-0.81	0.55	0.00	0.09	-0.77	0.53	-0.82	0.82	0.62	0.78	-0.84	0.52	-0.63	-0.66	-0.56	-0.37	-0.40	-0.64	0.02	0.41	-0.62	0.33	0.56	-0.78	-0.49	-0.62	-0.67	0.68
aridity	-0.19	0.07	0.24	0.36	0.11	0.12	-0.62	0.70	-0.06	0.42	-0.06	0.20	-0.34	0.01	0.08	-0.07	0.15	0.05	-0.48	0.49	-0.25	0.28	-0.71	-0.16	0.07	0.44	-0.08	0.12	0.34	0.27	-0.00	0.48
high_prec_freq	-0.52	-0.29	0.25	0.72	0.52	-0.23	-0.50	0.52	0.42	0.07	0.43	-0.37	-0.69	-0.48	0.54	-0.33	0.51	0.43	-0.06	0.58	-0.01	0.55	-0.61	-0.37	0.44	0.16	-0.34	0.48	0.39	0.59	0.39	-0.05
high_prec_dur	-0.02	-0.08	0.49	-0.10	-0.37	0.48	-0.40	0.54	-0.12	0.57	-0.12	0.54	0.14	0.26	-0.06	0.04	0.04	-0.41	-0.39	0.02	-0.64	0.06	-0.49	-0.33	-0.03	0.35	-0.27	-0.39	0.16	-0.22	-0.14	0.52
low_prec_freq	-0.52	-0.24	0.30	0.74	0.44	-0.13	-0.65	0.70	0.34	0.21	0.37	-0.21	-0.70	-0.40	0.49	-0.35	0.50	0.35	-0.23	0.64	-0.14	0.57	-0.78	-0.39	0.43	0.30	-0.37	0.41	0.47	0.57	0.32	0.15
low_prec_dur	-0.36	-0.28	0.56	0.31	0.02	0.27	-0.65	0.77	0.15	0.52	0.16	0.23	-0.31	-0.06	0.27	-0.17	0.34	-0.09	-0.36	0.34	-0.54	0.40	-0.73	-0.48	0.23	0.36	-0.45	-0.03	0.45	0.14	0.10	0.44
carbonate_rocks_frac	-0.22	-0.04	-0.14	0.22	0.25	-0.16	0.03	0.00	0.07	-0.17	0.13	-0.16	-0.19	-0.13	0.15	-0.11	0.07	0.18	0.11	0.24	0.18	0.09	-0.04	0.10	0.05	0.04	-0.01	0.22	0.26	0.23	0.05	-0.11
geol_permeability	-0.18	-0.13	0.01	0.15	0.09	-0.08	-0.19	0.17	0.15	0.04	0.16	-0.04	-0.18	-0.15	0.15	-0.33	0.23	0.00	-0.17	0.04	-0.16	0.21	-0.20	-0.17	0.23	-0.03	-0.25	0.04	0.08	0.07	0.06	0.00
soil_depth_pelletier	-0.25	-0.08	-0.18	0.37	0.41	-0.57	0.23	-0.21	0.23	-0.49	0.30	-0.32	-0.30	-0.32	0.29	-0.22	0.14	0.46	0.30	0.39	0.41	0.17	-0.01	0.15	0.18	0.09	0.01	0.45	0.15	0.49	0.33	-0.45
soil_depth_statsgo	-0.02	0.08	-0.35	0.07	0.19	-0.36	0.25	-0.31	0.05	-0.34	0.07	-0.20	-0.09	-0.13	0.04	-0.16	0.01	0.20	0.18	0.07	0.31	-0.02	0.16	0.19	0.04	-0.05	0.11	0.19	-0.09	0.16	0.06	-0.32
soil_porosity	-0.10	0.09	-0.16	0.09	0.21	-0.05	0.24	-0.18	-0.10	-0.23	0.00	-0.11	-0.00	0.05	0.02	0.22	-0.18	0.17	0.32	0.27	0.38	-0.15	0.14	0.26	-0.21	0.15	0.21	0.20	0.30	0.17	-0.04	-0.10
soil_conductivity	0.10	-0.04	-0.05	-0.10	-0.14	-0.14	0.04	-0.06	0.09	-0.04	0.01	0.05	0.06	-0.05	-0.02	-0.22	0.08	-0.07	-0.13	-0.20	-0.17	0.04	0.06	-0.03	0.13	-0.20	-0.06	-0.09	-0.32	-0.08	0.05	-0.10
max_water_content	-0.09	0.01	-0.29	0.16	0.30	-0.42	0.29	-0.34	0.14	-0.41	0.16	-0.33	-0.15	-0.22	0.14	-0.12	0.05	0.32	0.31	0.14	0.40	0.01	0.17	0.18	0.08	-0.07	0.10	0.31	-0.03	0.26	0.17	-0.43
sand_frac	0.17	-0.05	0.03	-0.14	-0.19	-0.06	-0.04	-0.02	0.09	0.07	-0.02	0.05	0.08	-0.05	-0.05	-0.19	0.10	-0.12	-0.19	-0.30	-0.25	0.05	0.04	-0.13	0.15	-0.27	-0.09	-0.16	-0.43	-0.16	0.06	-0.05
silt_frac	0.16	0.24	-0.19	-0.17	0.01	0.03	0.40	-0.36	-0.21	-0.20	-0.17	-0.02	0.22	0.19	-0.18	0.41	-0.32	0.05	0.34	-0.02	0.36	-0.34	0.34	0.35	-0.35	0.09	0.37	0.01	0.03	-0.06	-0.13	-0.14
clay_frac	-0.42	-0.27	0.20	0.50	0.44	-0.16	-0.24	0.23	0.32	-0.07	0.39	-0.36	-0.45	-0.37	0.42	-0.17	0.33	0.38	0.15	0.41	0.16	0.37	-0.32	-0.25	0.30	0.18	-0.28	0.40	0.41	0.43	0.31	-0.13
elev_mean	0.64	0.64	0.05	-0.42	-0.73	0.66	-0.40	0.47	-0.63	0.77	-0.73	0.79	0.45	0.66	-0.68	0.40	-0.39	-0.63	-0.82	-0.31	-0.59	-0.39	-0.31	0.13	-0.43	0.40	0.35	-0.72	-0.39	-0.53	-0.56	0.83
slope_mean	0.59	0.35	0.16	-0.67	-0.83	0.74	-0.14	0.21	-0.51	0.66	-0.59	0.74	0.62	0.64	-0.62	0.37	-0.38	-0.78	-0.54	-0.59	-0.67	-0.42	0.03	0.01	-0.39	0.12	0.15	-0.85	-0.38	-0.75	-0.55	0.66
area_gages2	-0.23	-0.10	-0.02	0.22	0.24	-0.18	-0.01	0.03	0.12	-0.18	0.19	-0.15	-0.22	-0.16	0.19	-0.14	0.11	0.21	0.11	0.27	0.15	0.13	-0.08	0.01	0.10	0.08	-0.08	0.22	0.24	0.24	0.13	-0.11
frac_forest	0.42	0.05	-0.03	-0.58	-0.43	0.28	0.25	-0.33	-0.14	0.12	-0.28	0.16	0.50	0.25	-0.32	0.19	-0.20	-0.42	-0.01	-0.71	-0.21	-0.32	0.46	-0.02	-0.18	-0.45	0.09	-0.45	-0.53	-0.59	-0.21	-0.01
lai_max	0.13	-0.18	-0.11	-0.27	0.01	-0.10	0.41	-0.54	0.16	-0.31	0.09	-0.33	0.20	-0.12	0.02	0.02	-0.02	0.01	0.38	-0.48	0.21	-0.13	0.60	-0.02	0.03	-0.61	0.01	-0.00	-0.36	-0.23	0.09	-0.43
lai_diff	0.17	-0.04	-0.28	-0.25	0.07	-0.14	0.47	-0.61	0.08	-0.39	0.01	-0.37	0.17	-0.09	-0.05	0.10	-0.12	0.08	0.41	-0.40	0.37	-0.21	0.68	0.16	-0.07	-0.60	0.19	0.07	-0.34	-0.17	0.01	-0.43
gvf_max	0.13	-0.14	-0.18	-0.28	0.02	-0.12	0.53	-0.64	0.13	-0.40	0.08	-0.33	0.25	-0.09	0.01	0.07	-0.08	0.02	0.48	-0.44	0.30	-0.20	0.67	0.08	-0.04	-0.56	0.07	0.01	-0.32	-0.22	0.06	-0.50
gvf_diff	0.17	0.28	-0.54	-0.08	0.19	-0.23	0.55	-0.59	-0.14	-0.50	-0.12	-0.32	0.07	0.02	-0.14	0.27	-0.33	0.21	0.45	-0.01	0.66	-0.36	0.63	0.53	-0.30	-0.25	0.52	0.21	-0.11	0.06	-0.13	-0.36

- Representation learning model generates embeddings that highly correlates with known physical characteristics

Summary and Future Work

- Random vectors can be used for source modulation when basin characteristics are not available
    - Show robust performance even under data sparsity and different modulation strategies

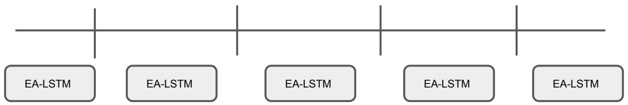
- Representation learning can be used to address the issue of missing or uncertain characteristics by learning them from data
- Self-supervised representation learning based on contrastive loss show promise in identifying unknown characteristics

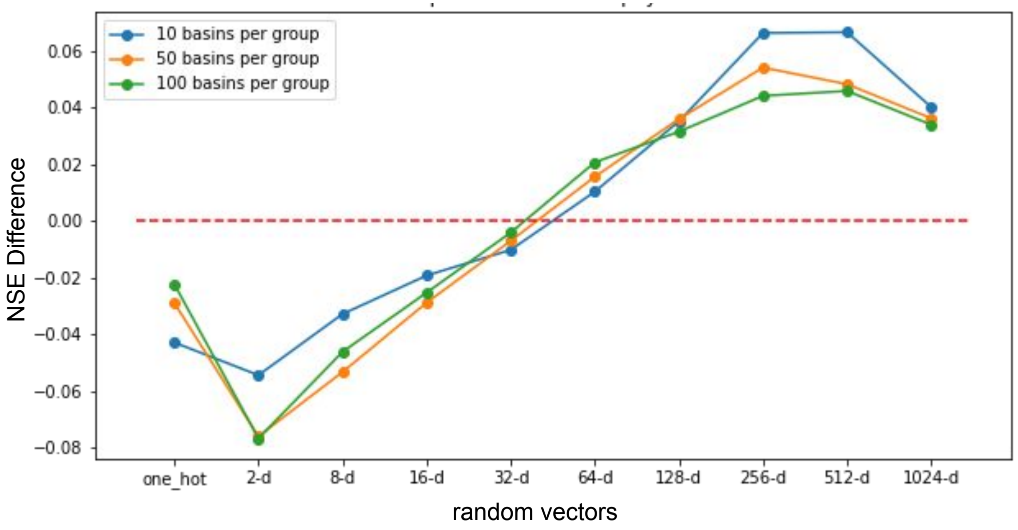
Future Work

- Comparison of Source Aware Modulation with traditional Meta-Learning (e.g. MAML)
- New ways of representation learning

Backup slides

Data Sparsity: Impact of number of sources using EA-LSTM





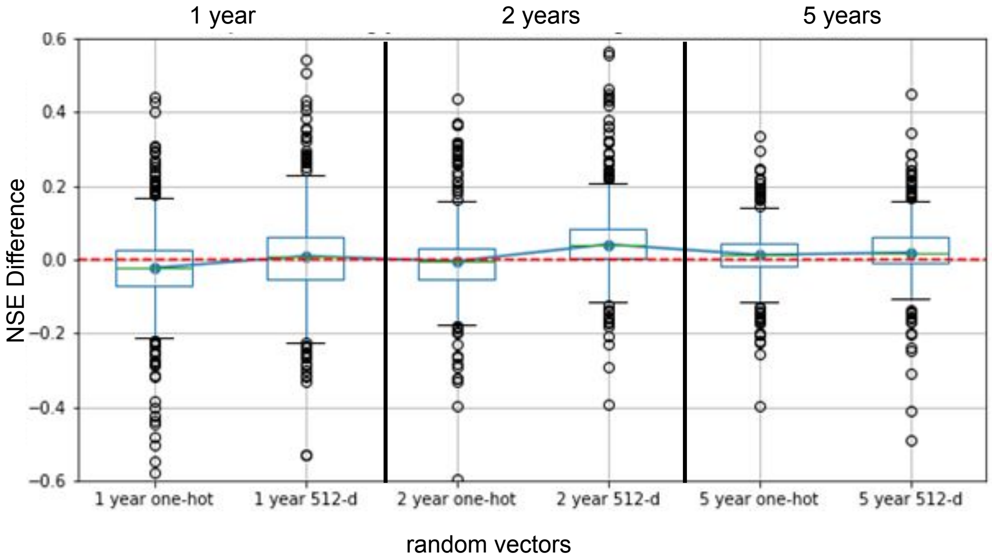
X-axis represents models that use randomly assigned vectors

Y-axis represents increase in performance compared to using 27-d physical characteristics

- NSE using random vectors - NSE using physical characteristics
- Box plot of 531 values (corresponding to 531 basins)
- Positive values suggest random vectors perform better than physical characteristics

- Even with very few sources, modulation can be done using random vectors
- Random vectors with higher dimensions perform slightly better than physical characteristics
- One-hot vector performs poorly compared to random vectors

Data Sparsity: Impact of training data size using EA-LSTM



X-axis represents models that use random vectors

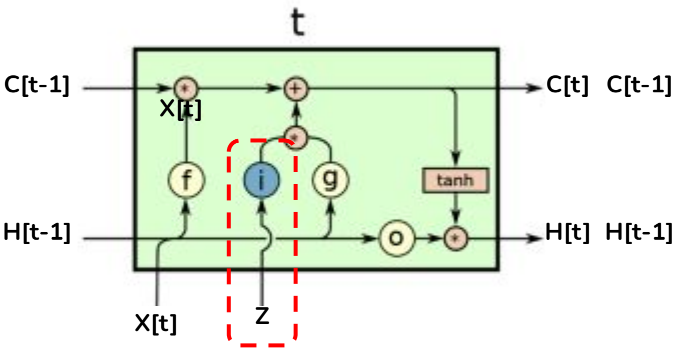
Y-axis represents difference in performance compared to using 27-d physical characteristics

- NSE using random vectors - NSE using physical characteristics
- Box plot of 531 values (corresponding to 531 basins)
- Positive values suggest that random vectors perform better than physical characteristics

- Random vectors show robust performance even with small number of years available for training

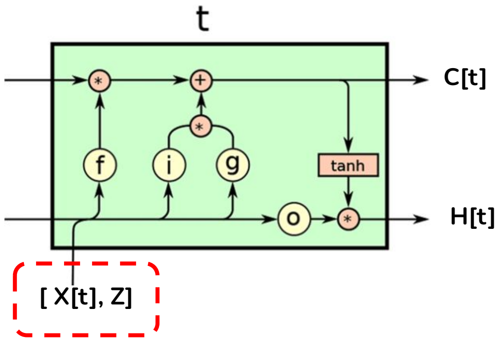
Impact of Modulation Strategy

- EA-LSTM is one of the ways of modulation



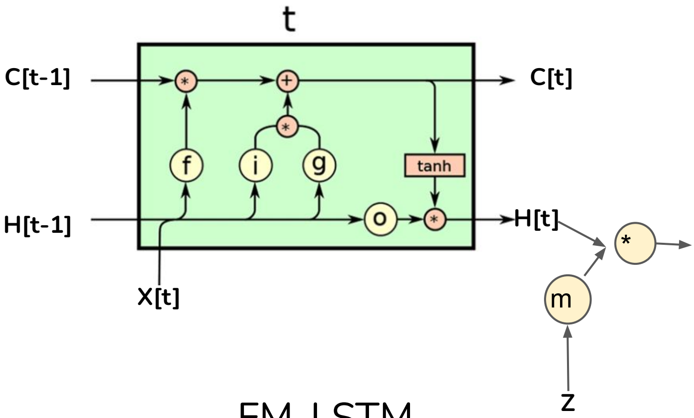
EA-LSTM

EA-LSTM uses input gate as a modulator driven by static source characteristics to modulate the dynamic cell state



CT-LSTM

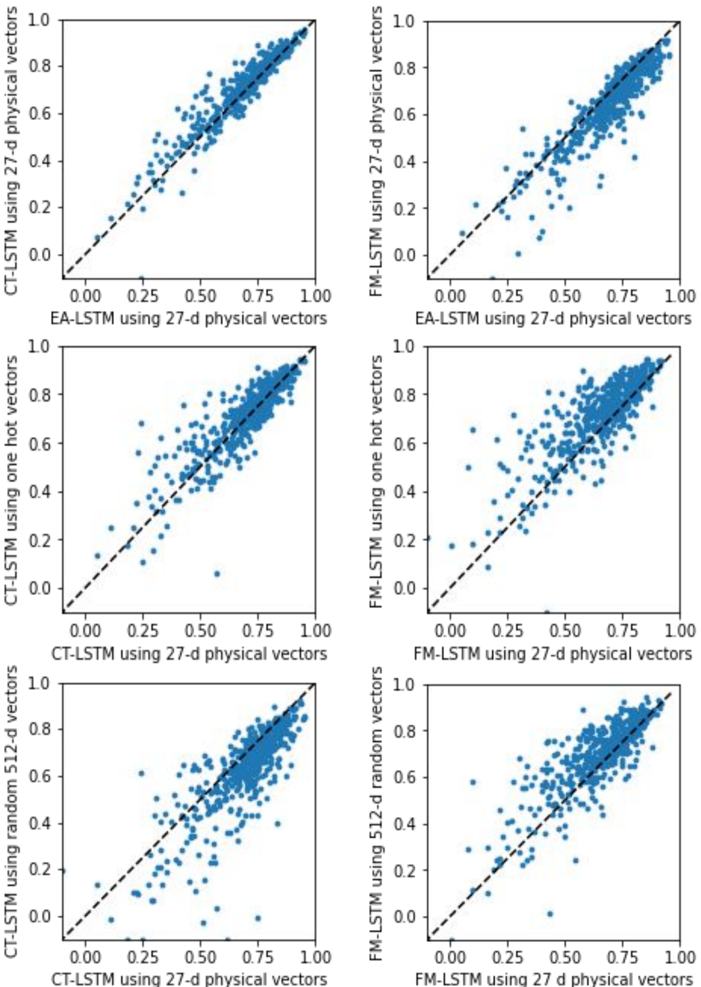
CT-LSTM concatenates source characteristics with dynamic input variables and uses a vanilla LSTM



FM-LSTM

FM-LSTM uses a separate modulation gate to update hidden features generated by a vanilla LSTM

Impact of Modulation Strategy

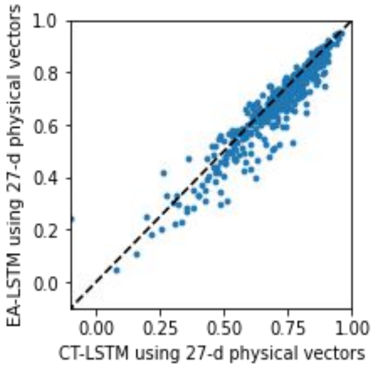


- CT-LSTM > EA-LSTM > FM-LSTM when physical vectors are used

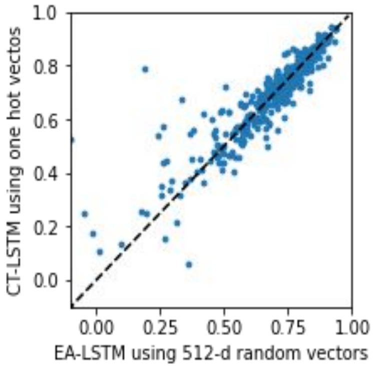
- One-hot vectors show robust performance using both CT-LSTM and FM-LSTM as modulation strategies
  - Performance significantly better than physical vectors when FM-LSTM is used

- Random vectors with high dimensions perform poorly when CT-LSTM is used

Which modulation strategy is better ?

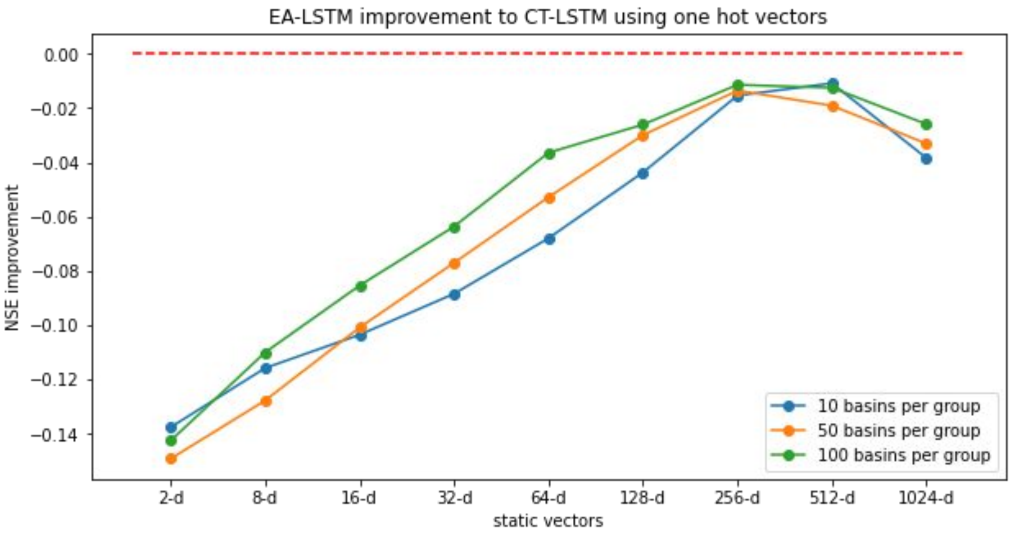


- CT-LSTM performs better than EA-LSTM when physical vectors are available but at the expense of interpretability



- Both EA-LSTM (with random vectors) and CT-LSTM (with one-hot vectors) show comparable performance

Which modulation strategy is better under data sparsity?



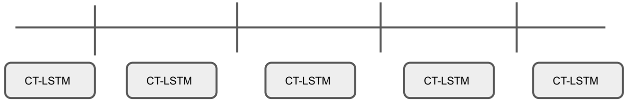
X-axis represents EA-LSTM models that use random vectors

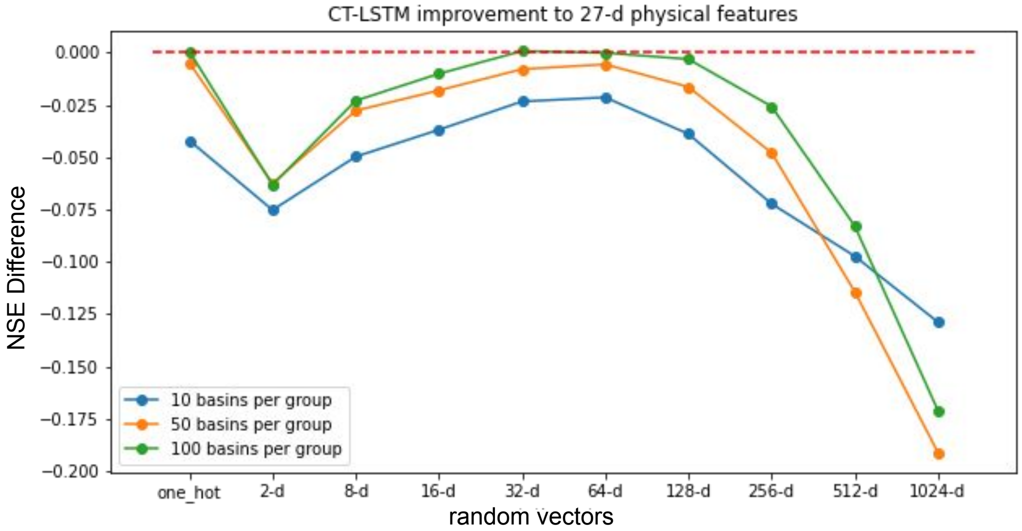
Y-axis represents increase in performance compared to using CT-LSTM model with one-hot vectors

- NSE using EA-LSTM with random vectors - NSE using CT-LSTM with uncorrelated one-hot vectors
- Box plot of 531 values (corresponding to 531 basins)
- Positive values suggest EA-LSTM perform better than CT-LSTM

- CT-LSTM with one-hot vectors perform slightly better even in data sparse scenario

Data Sparsity: Impact of number of sources on CT-LSTM





X-axis represents models that use non-informative source characteristics

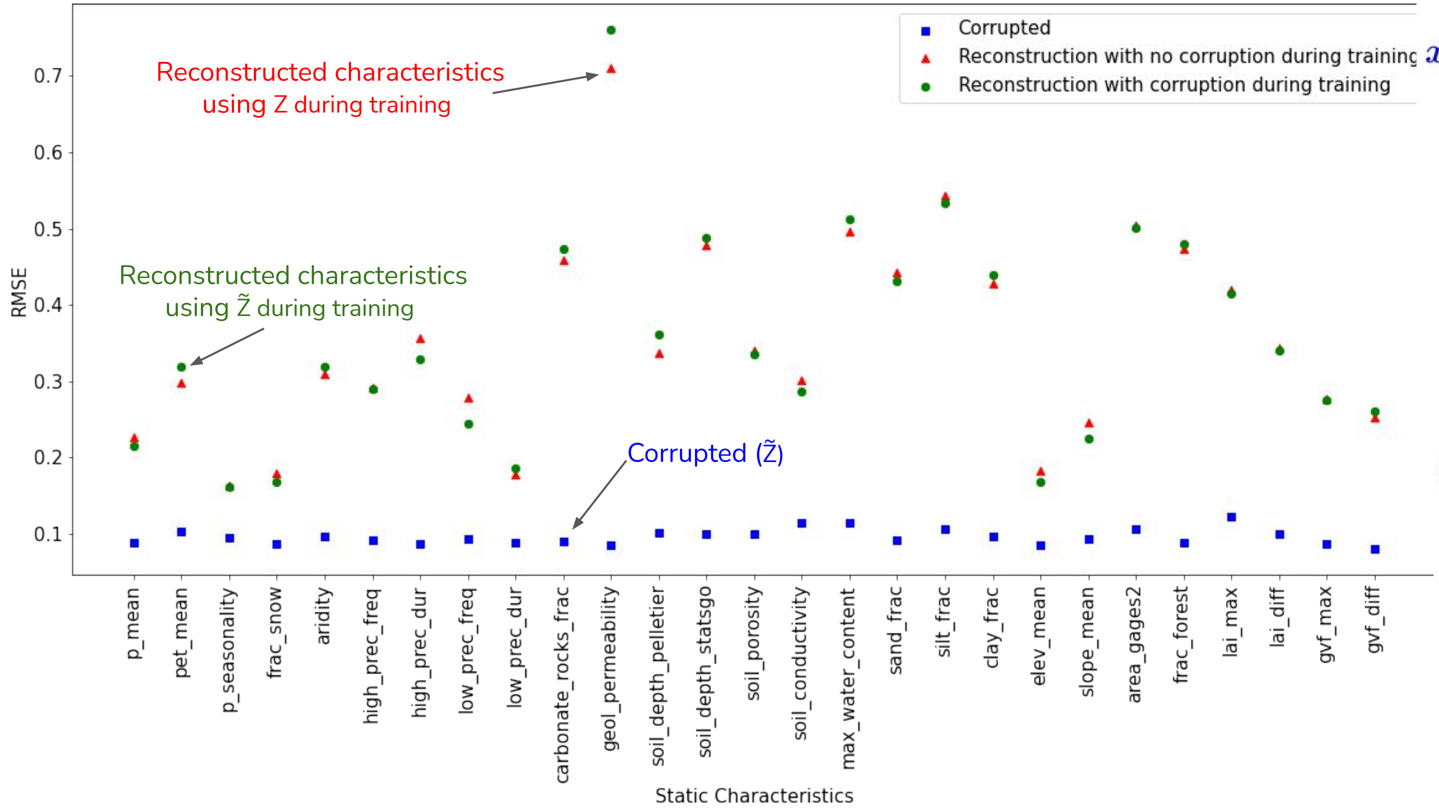
Y-axis represents increase in performance compared to using physical characteristics

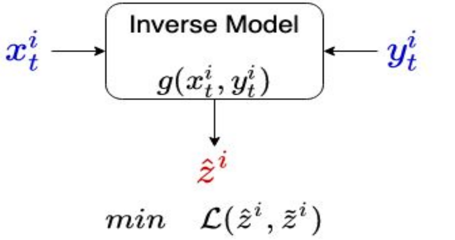
- NSE using random characteristics - NSE using physical characteristics
- Box plot of 531 values (corresponding to 531 basins)
- Positive values suggest that random vectors perform better than physical characteristics

- Physical characteristics perform slightly better than random vectors
- Sweet spot occurs at much lower dimensions in case of CT-LSTM
- One-hot vector performs comparable to random vectors unlike EA-LSTM

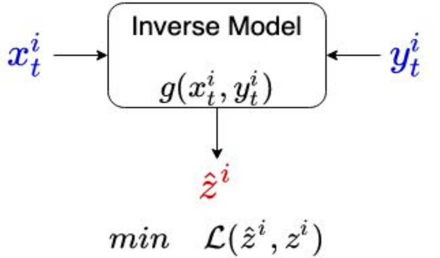
Impact of measurement uncertainty in basin characteristics

- 0.1 std. deviation in 10 % values





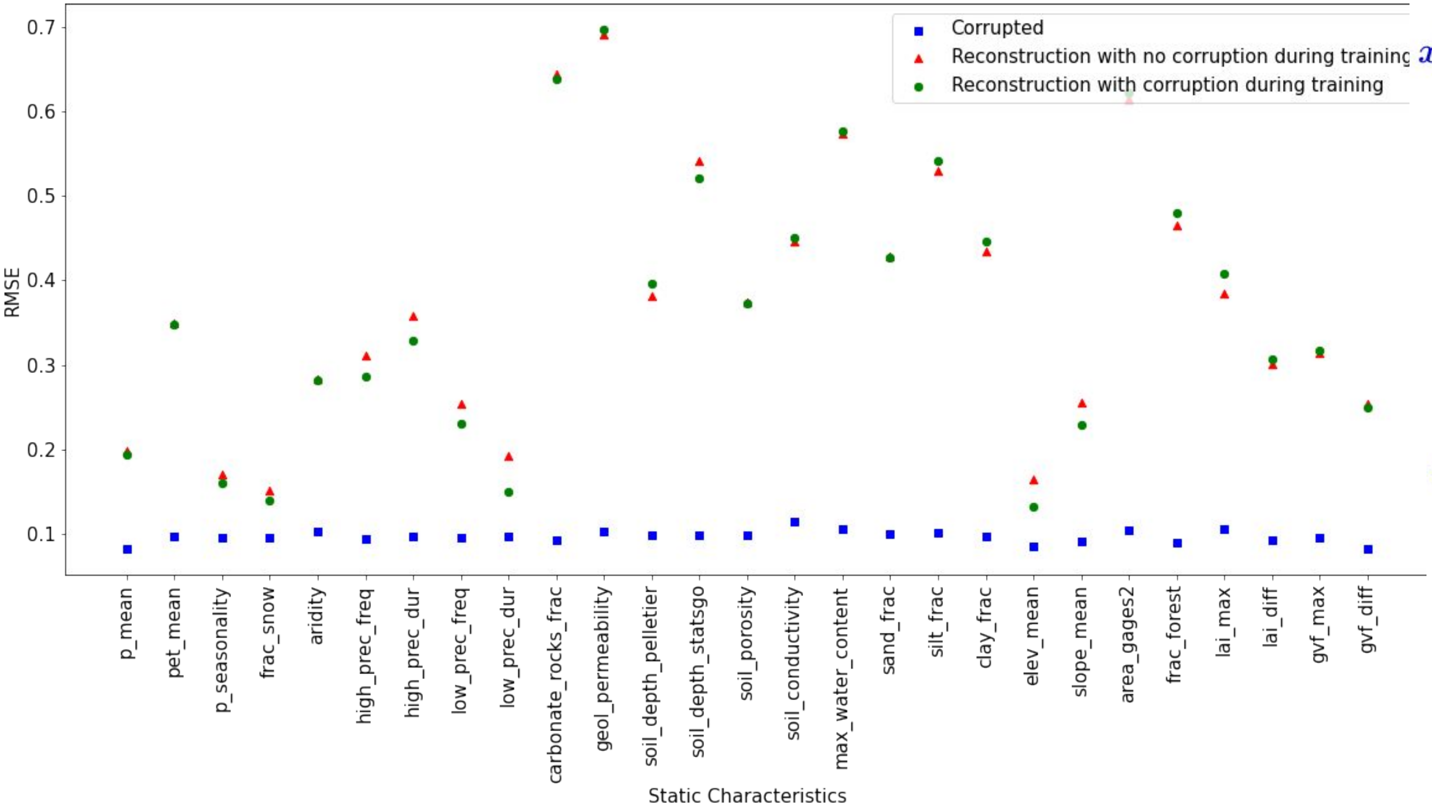
Trained using corrupted characteristics

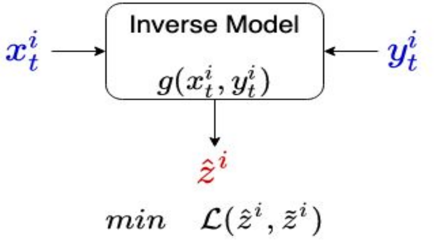


Trained using true characteristics

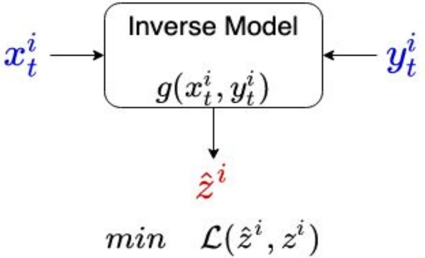
Impact of measurement uncertainty in basin characteristics

- 0.1 std. deviation in 20 % values





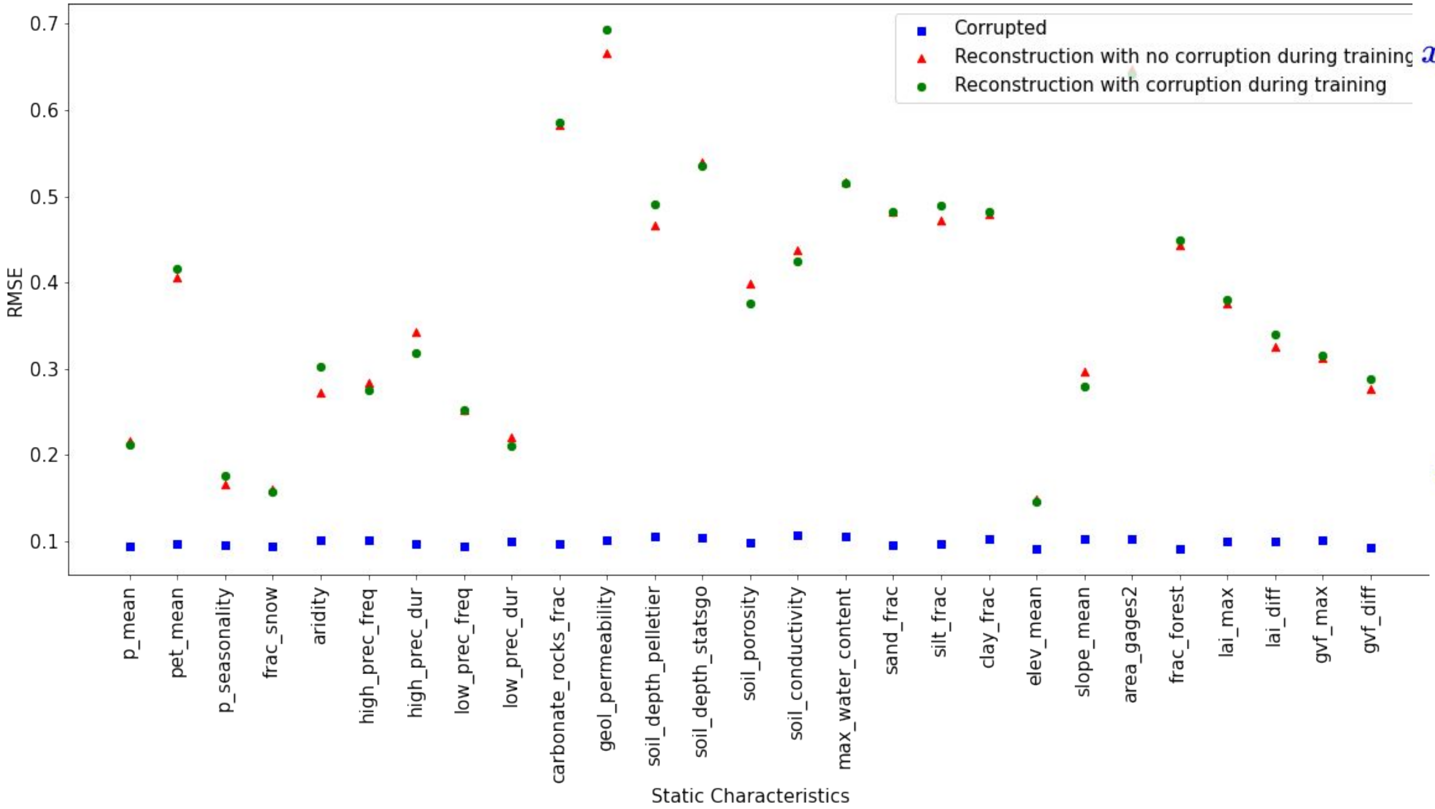
Trained using corrupted characteristics

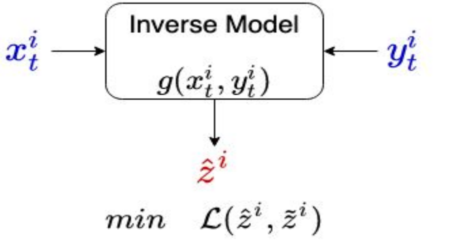


Trained using true characteristics

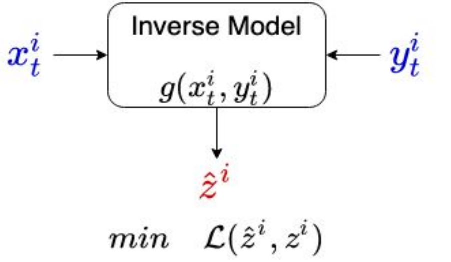
Impact of measurement uncertainty in basin characteristics

- 0.1 std. deviation in 50 % values





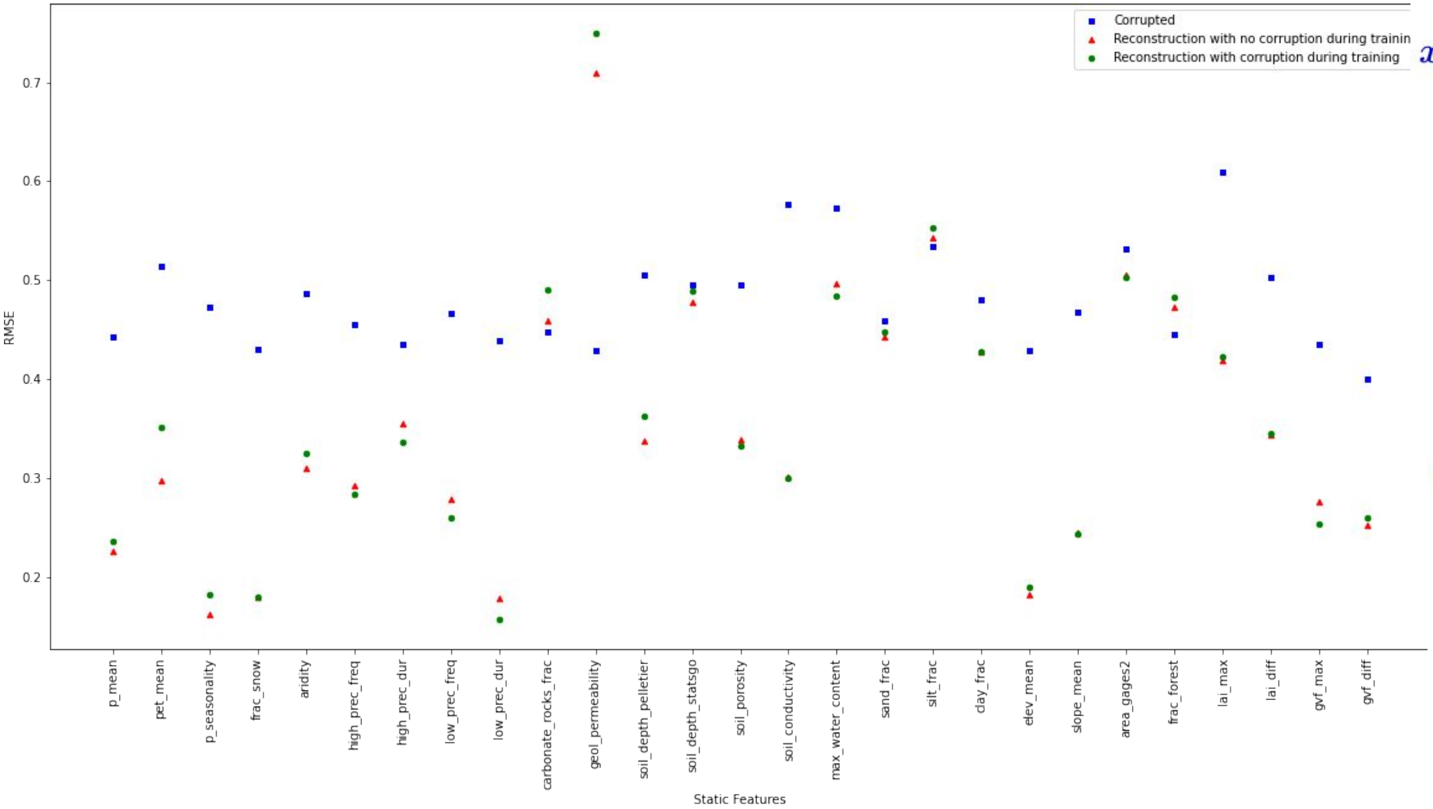
Trained using corrupted characteristics

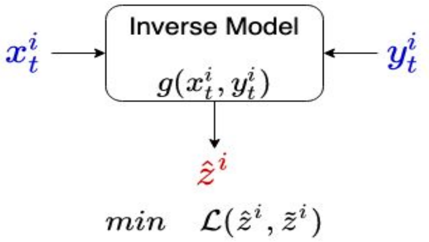


Trained using true characteristics

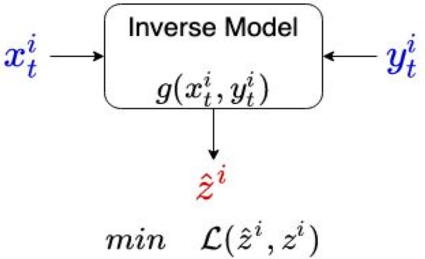
Impact of measurement uncertainty in basin characteristics

- 0.5 std. deviation in 10 % values





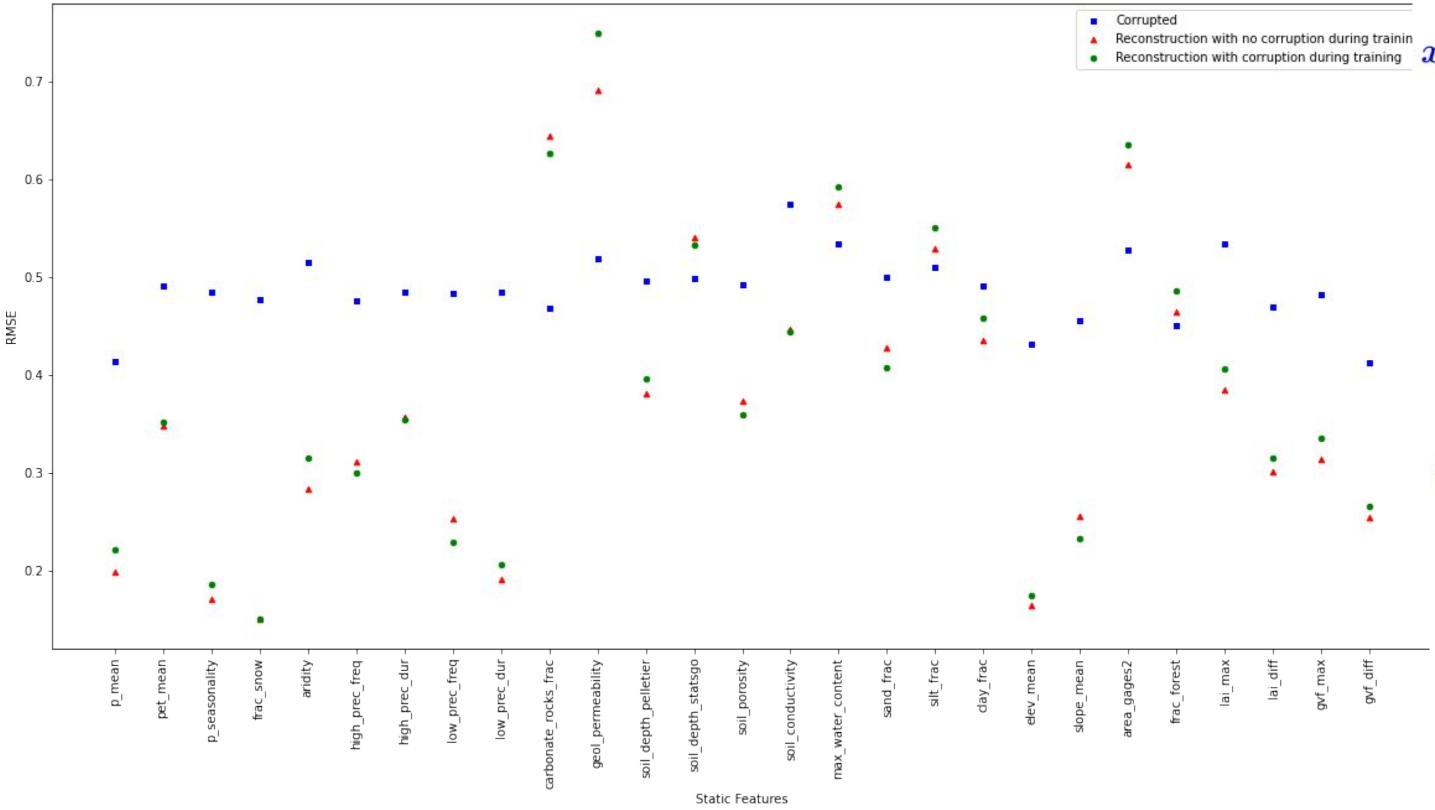
Trained using corrupted characteristics

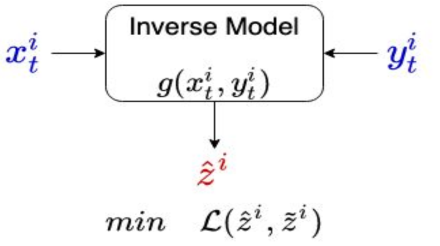


Trained using true characteristics

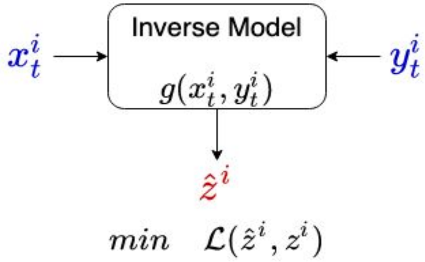
Impact of measurement uncertainty in basin characteristics

- 0.5 std. deviation in 20 % values





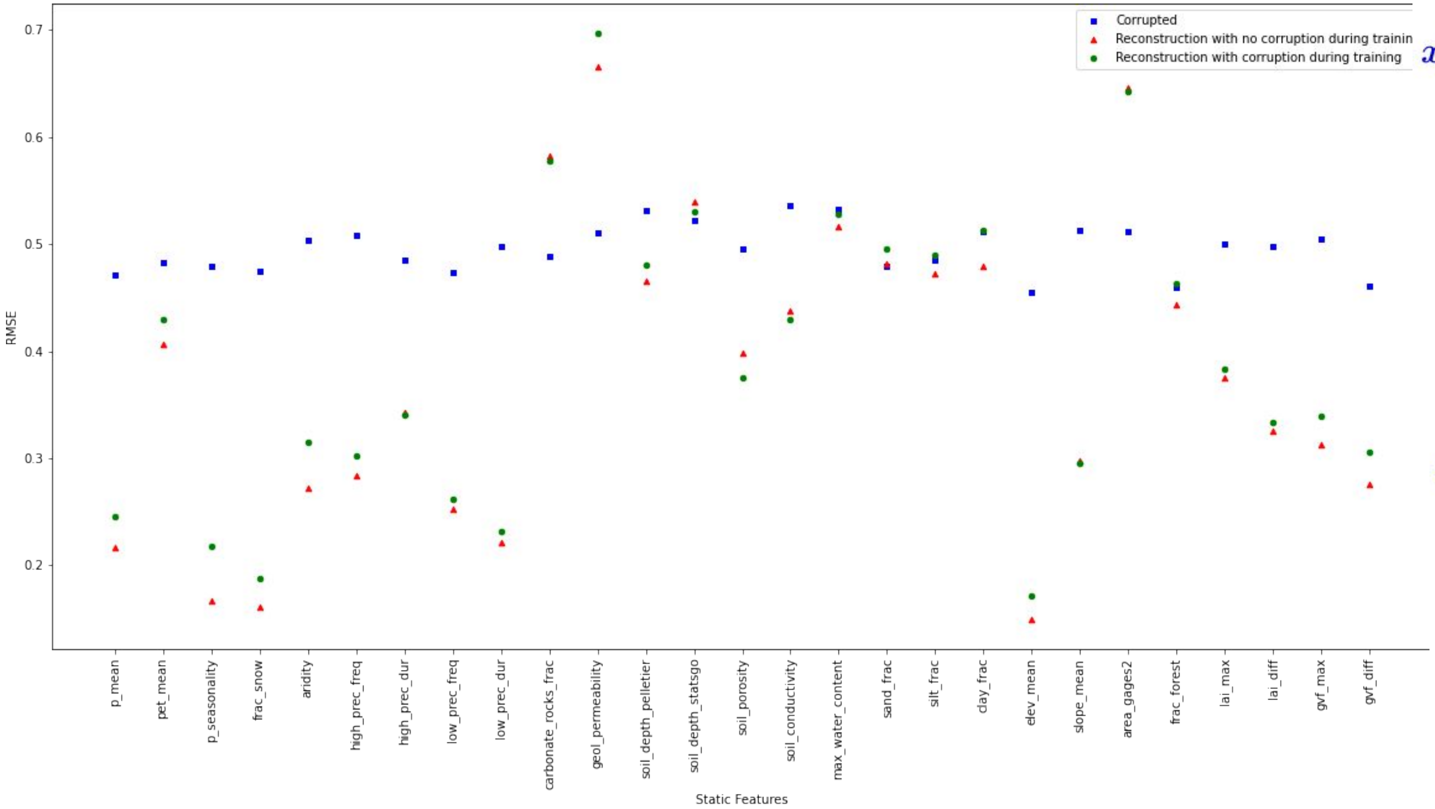
Trained using corrupted characteristics

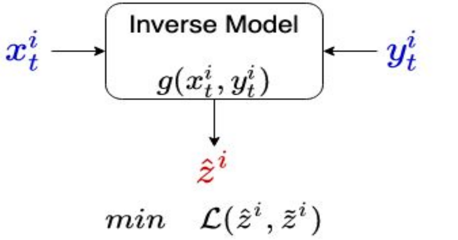


Trained using true characteristics

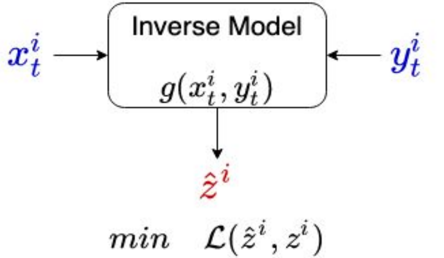
Impact of measurement uncertainty in basin characteristics

- 0.5 std. deviation in 50 % values





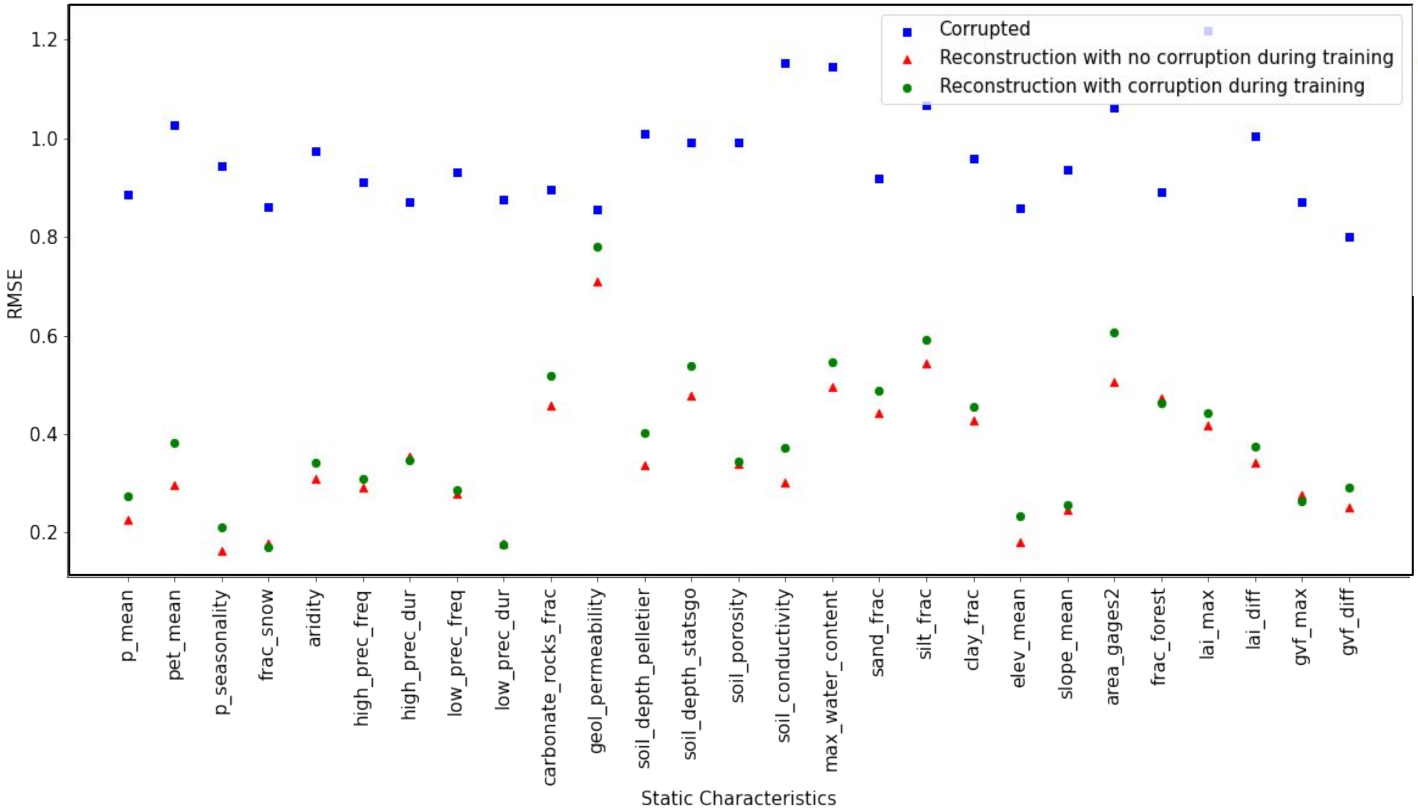
Trained using corrupted characteristics

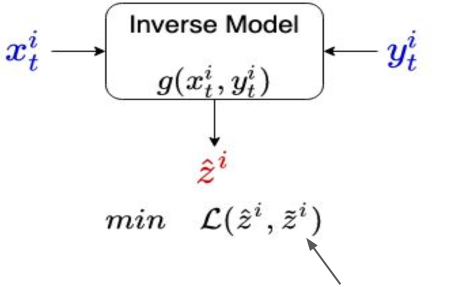


Trained using true characteristics

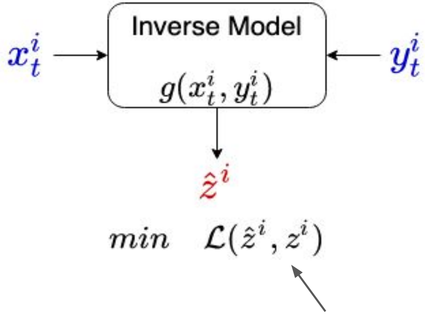
Impact of measurement uncertainty in basin characteristics

- 1 std. deviation in 10 % values





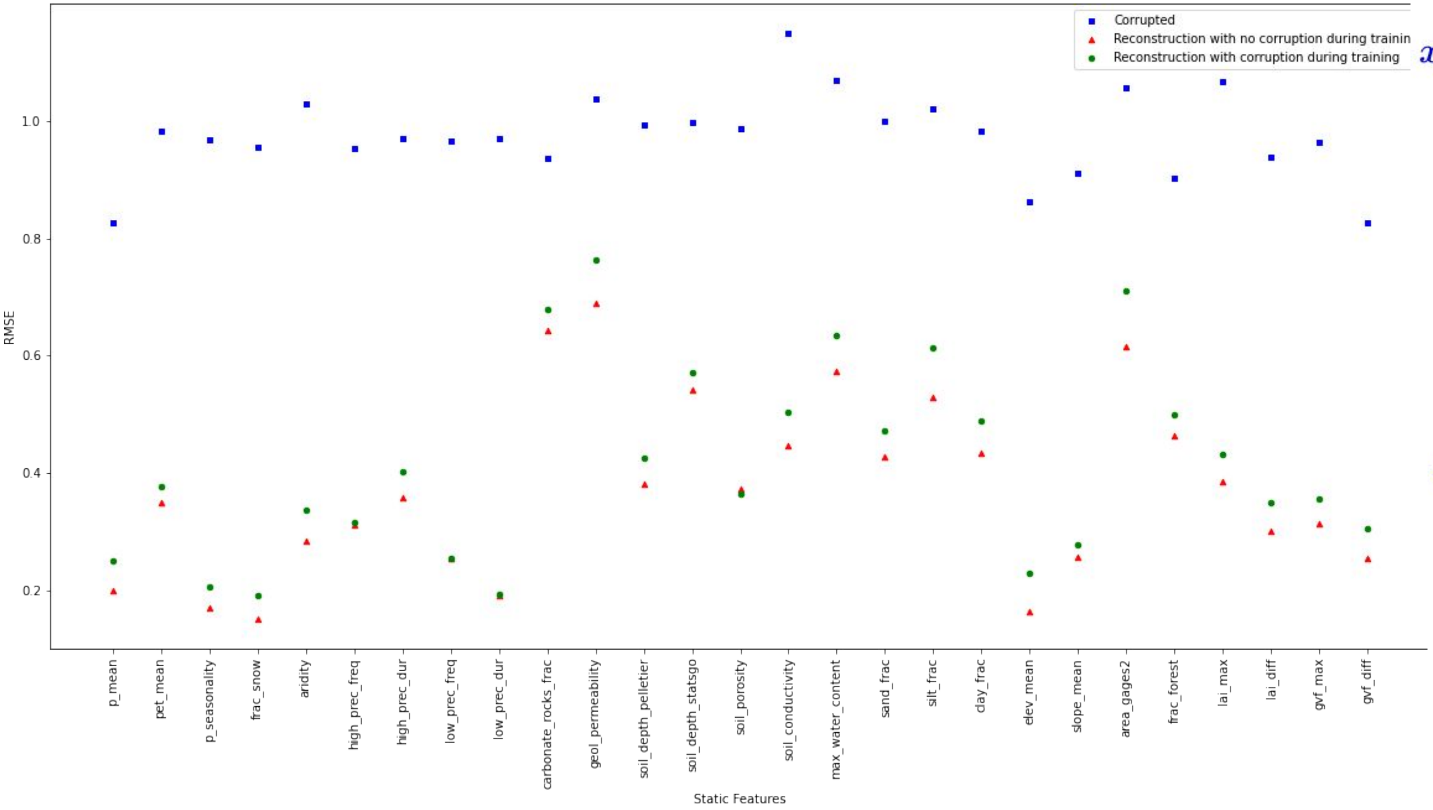
Trained using corrupted characteristics

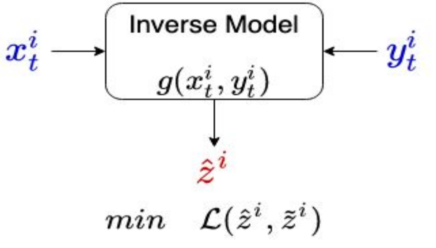


Trained using true characteristics

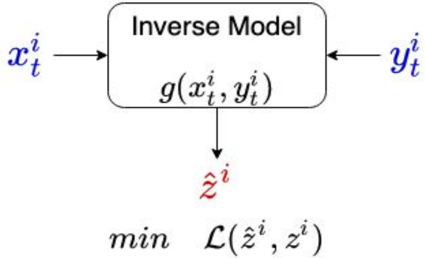
Impact of measurement uncertainty in basin characteristics

- 1 std. deviation in 20 % values





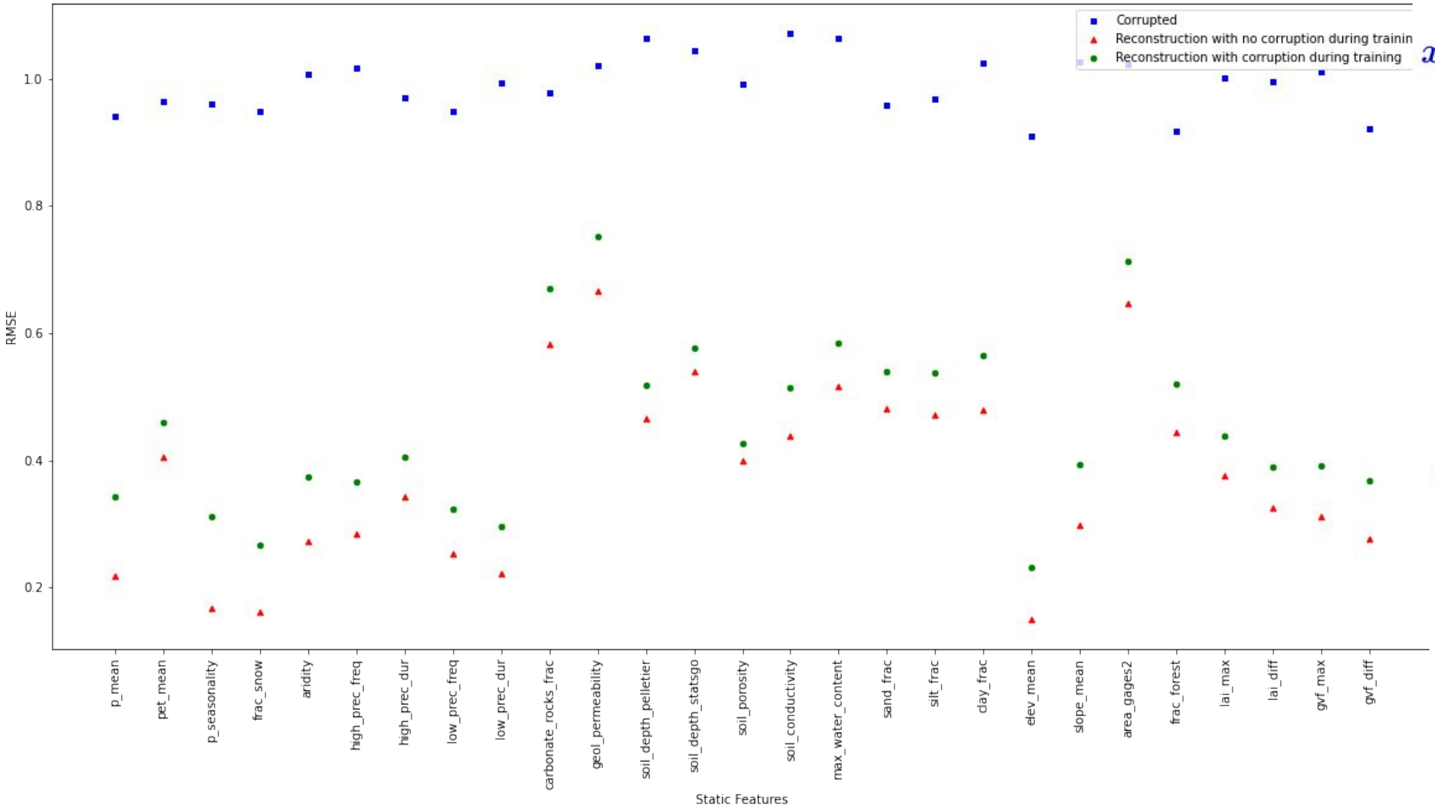
Trained using corrupted characteristics

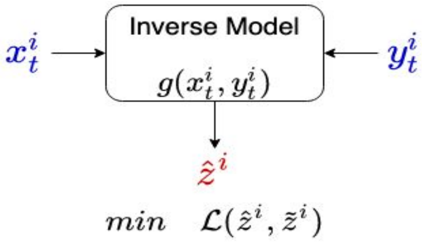


Trained using true characteristics

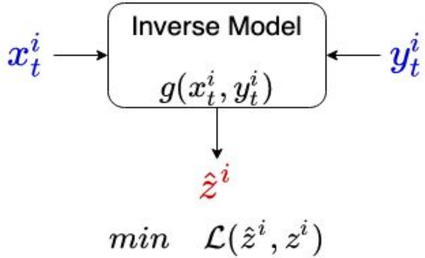
Impact of measurement uncertainty in basin characteristics

- 1 std. deviation in 50 % values





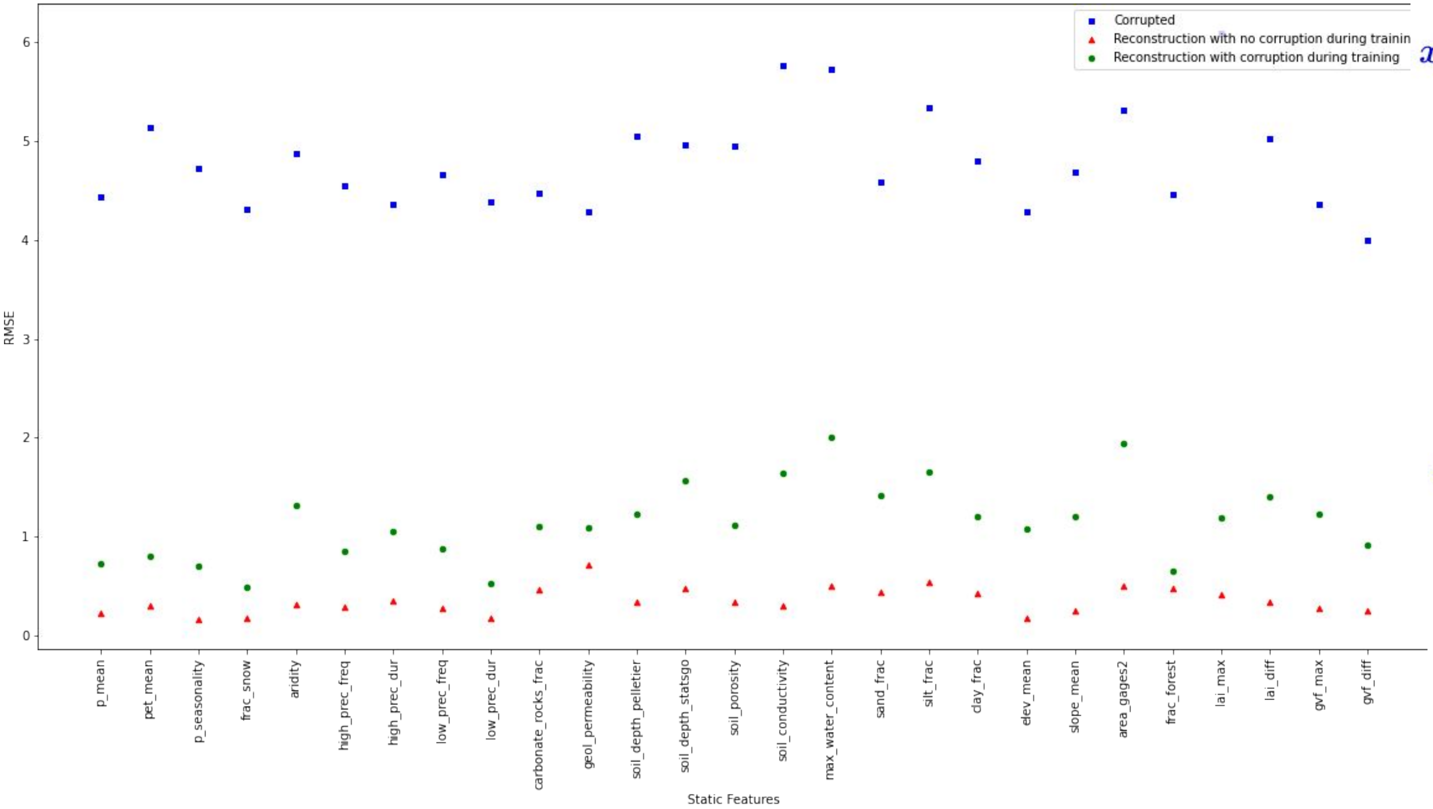
Trained using corrupted characteristics

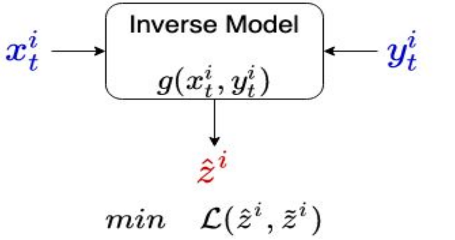


Trained using true characteristics

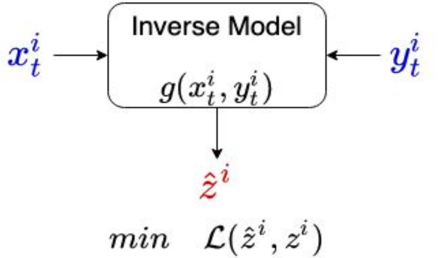
Impact of measurement uncertainty in basin characteristics

- 5 std. deviation in 10 % values





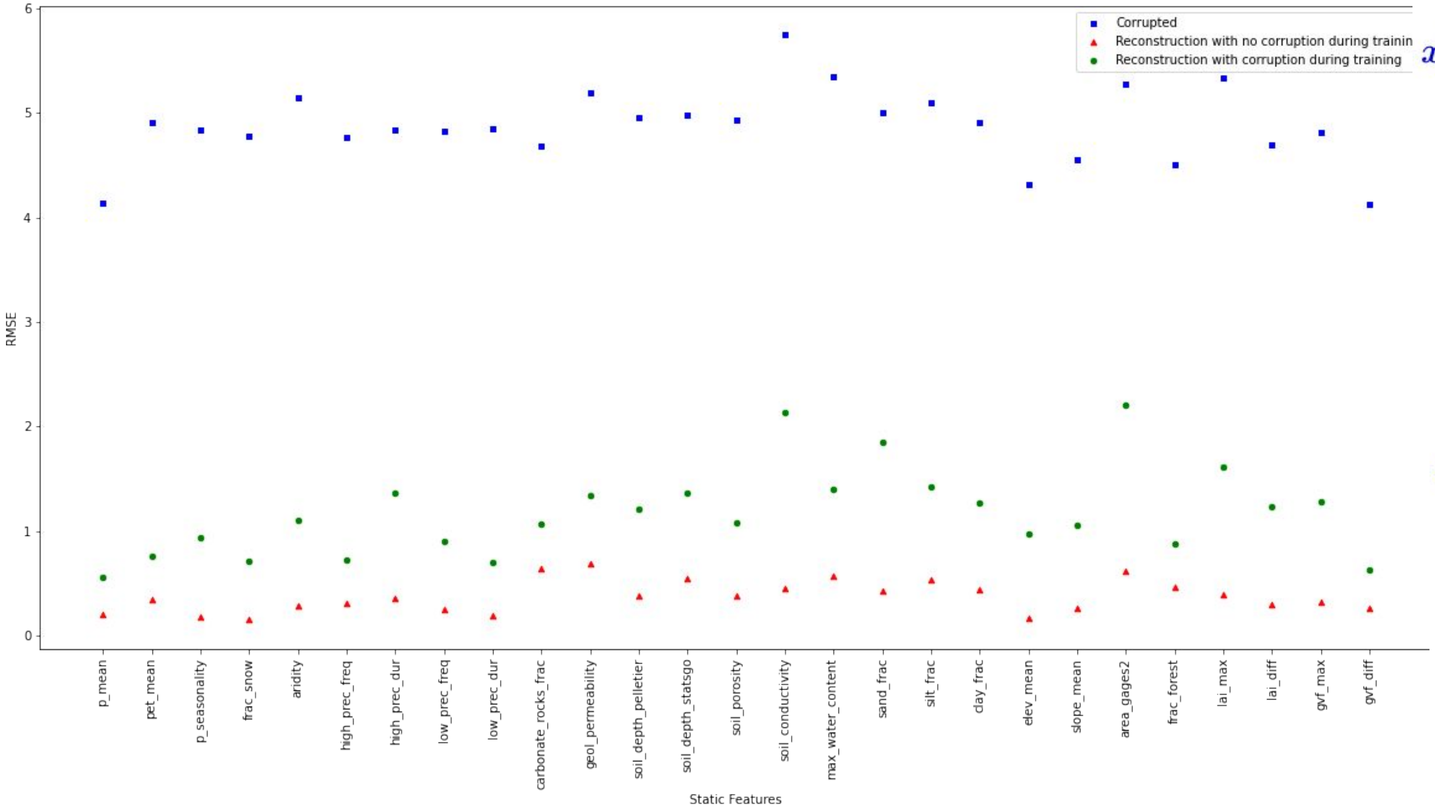
Trained using corrupted characteristics

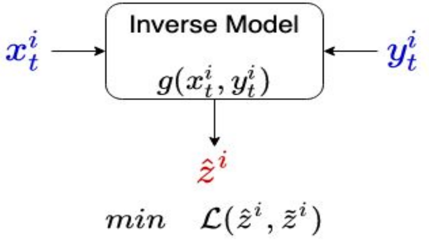


Trained using true characteristics

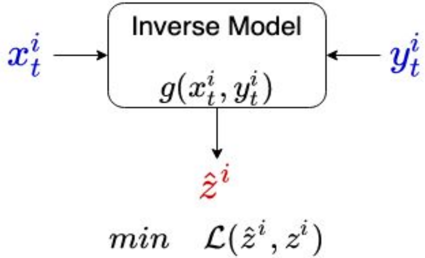
Impact of measurement uncertainty in basin characteristics

- 5 std. deviation in 20 % values





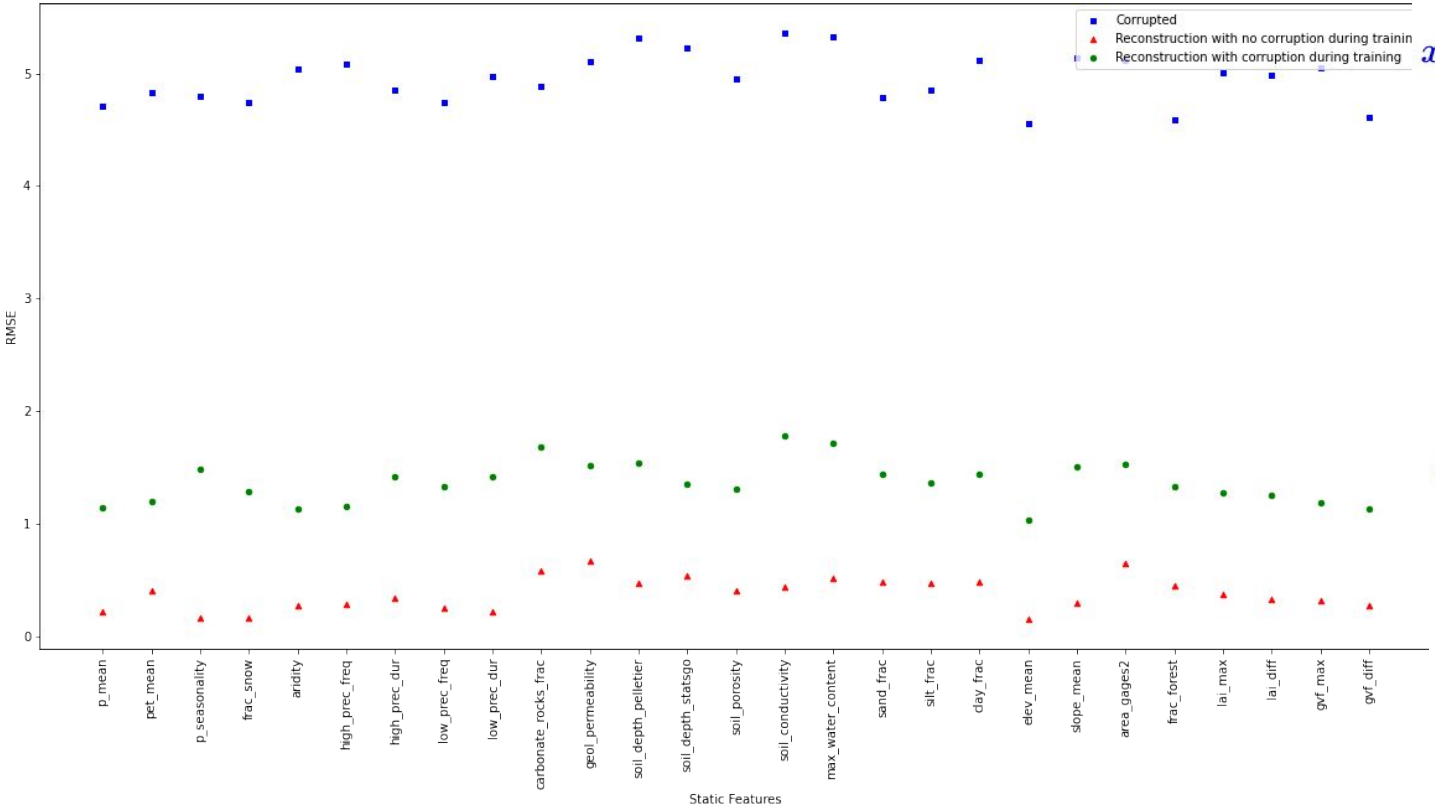
Trained using corrupted characteristics

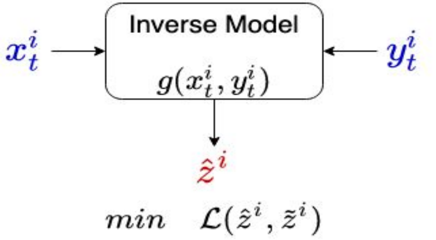


Trained using true characteristics

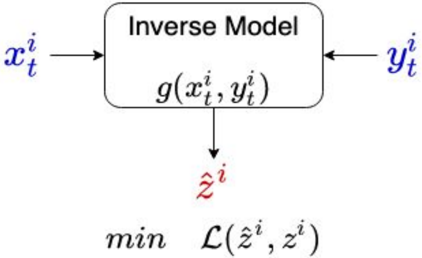
Impact of measurement uncertainty in basin characteristics

- 5 std. deviation in 50 % values



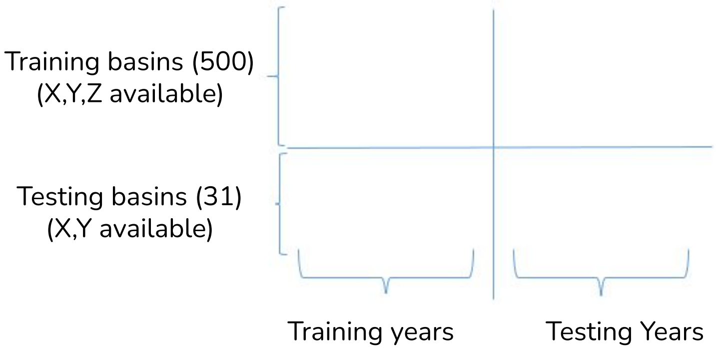


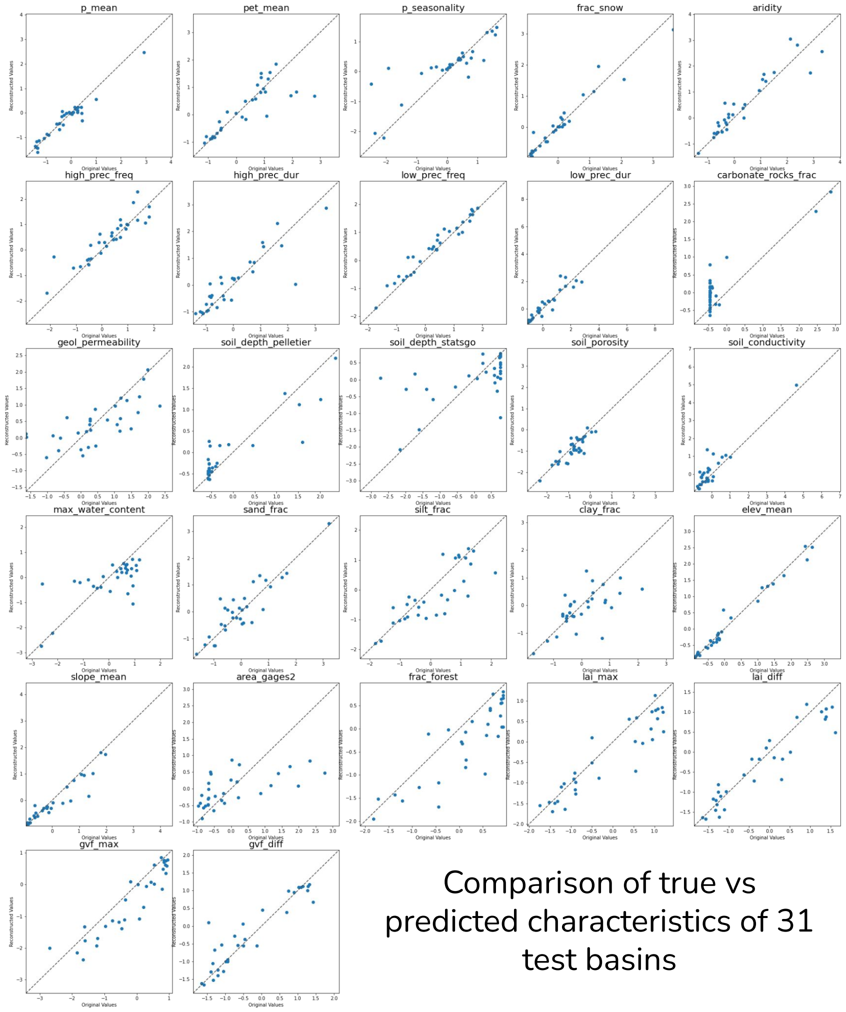
Trained using corrupted characteristics



Trained using true characteristics

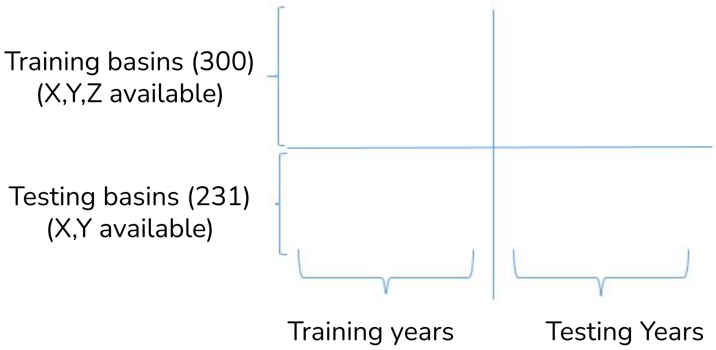
Impute missing characteristics

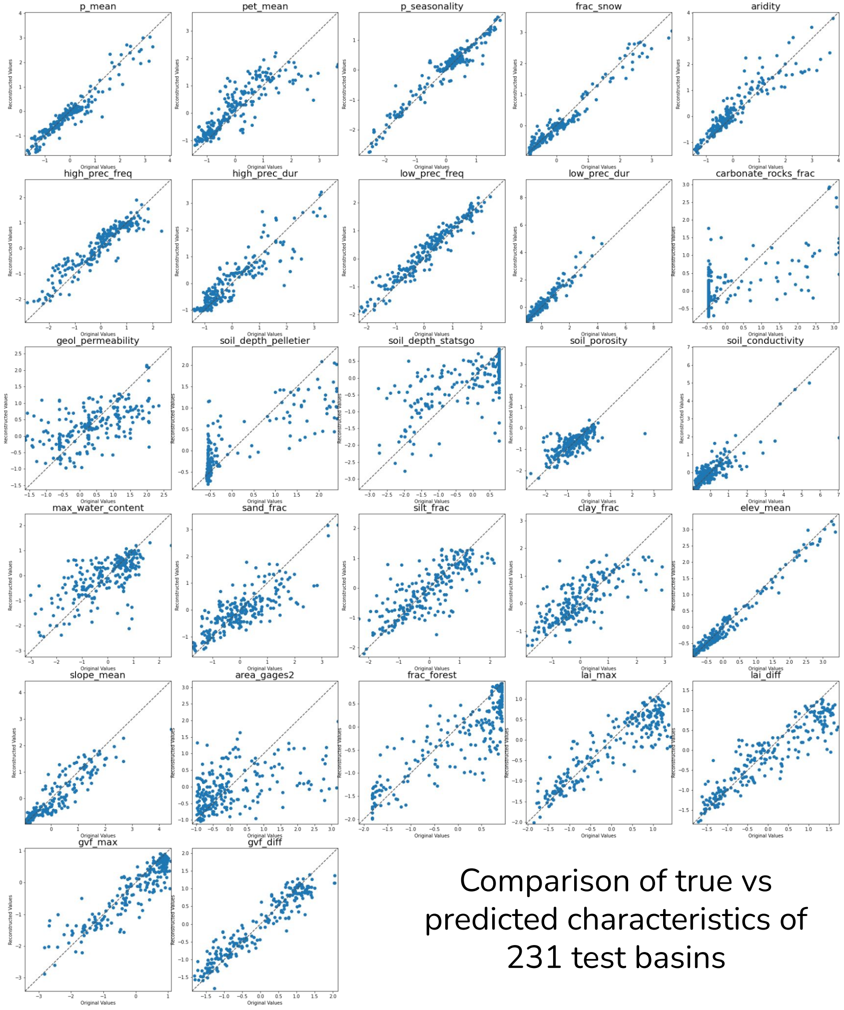




Comparison of true vs predicted characteristics of 31 test basins

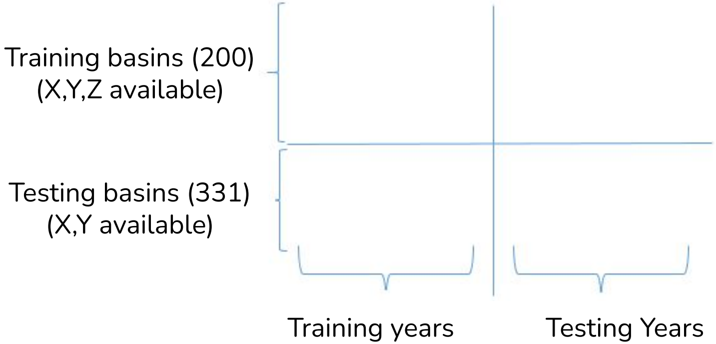
Impute missing characteristics

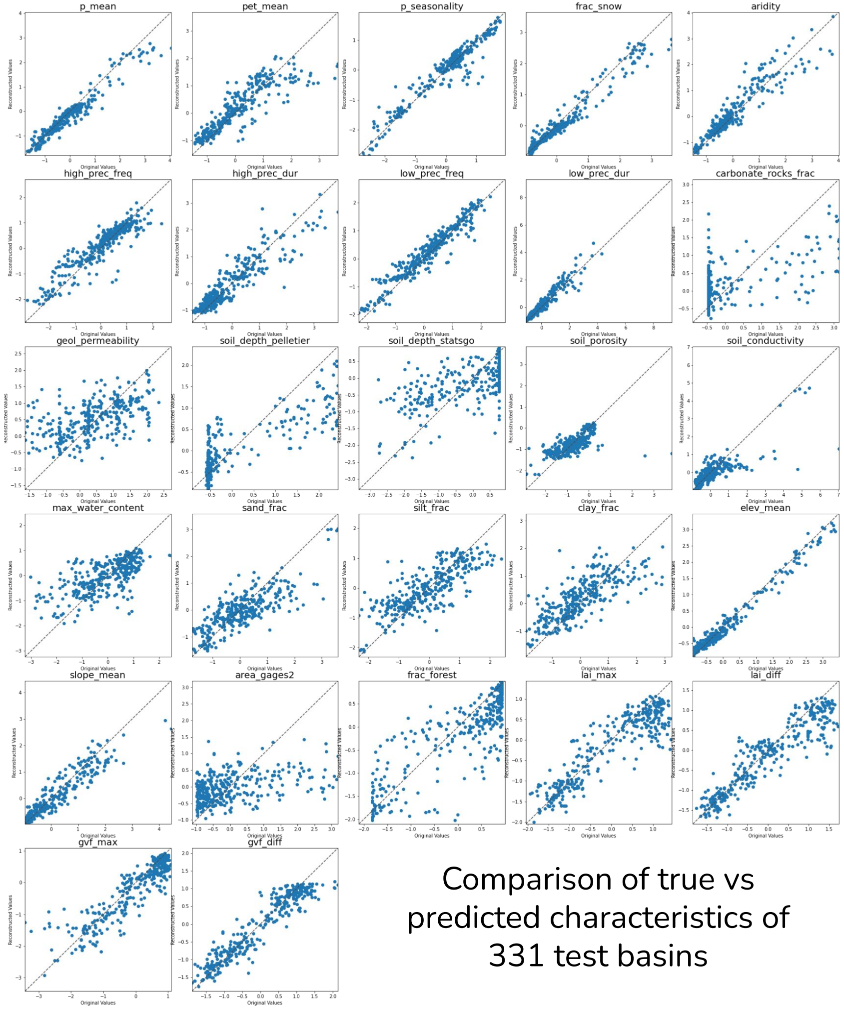




Comparison of true vs predicted characteristics of 231 test basins

Impute missing characteristics





Comparison of true vs predicted characteristics of 331 test basins